arXiv: math.PR/0000000

DETECTION OF BISTABILITY IN PHASE SPACE OF A REAL GALAXY, USING A NEW NON-PARAMETRIC BAYESIAN TEST OF HYPOTHESIS

BY DALIA CHAKRABARTY[*,†,‡,§],,

University of Warwick[‡] and University of Leicester[§]

In lieu of direct detection of dark matter, estimation of the distribution of the gravitational mass in distant galaxies is of crucial importance in Astrophysics. Typically, such estimation is performed using small samples of noisy, partially missing measurements - only some of the three components of the velocity and location vectors of individual particles that live in the galaxy are measurable. Such limitations of the available data in turn demands that simplifying model assumptions be undertaken. Thus, assuming that the phase space of a galaxy manifests simple symmetries - such as isotropy - allows for the learning of the density of the gravitational mass in galaxies. This is equivalent to assuming that the phase space pdf from which the velocity and location vectors of galactic particles are sampled from, is an isotropic function of these vectors.

We present a new non-parametric test of hypothesis that tests for relative support in two or more measured data sets of disparate sizes, for the undertaken model assumption that a given set of galactic particle data is sampled from an isotropic phase space pdfs. This test is designed to work in the context of Bayesian non-parametric, multimodal inference in a high-dimensional state space. In fact, the different models that are being compared are characterised by differential dimensionalities of the model parameter vectors. In addition, there is little prior information available about the unknown parameters, suggesting uninformative priors on the parameters in the different models. The problem of model parameter vectors of distinct dimensionalities and the difficulties of computing Bayes factors in this context, are circumvented in this test. The test works by identifying the subspace (of the system phase space) that is populated by those model parameter vectors, the posterior probability density of which exceed the maximal posterior density achieved under the null. The complement of the probability of the null given a data set is then the integral of the posterior probability density of the model parameters over this sub-space. This integral is the probability that the model parameter lives in this subspace; in implementational terms this probability is approximated as the fraction of the model parameters that live in this subspace. We illustrate applications of this test with two independent particle data sets in simulated as well as in a real galactic system.

The dynamical implications of the results of application to the real galaxy is indicated to be the residence of the observed particle samples in disjoint volumes of the galactic phase space. This result is used to suggest the serious risk borne in attempts at learning of gravitational mass density of galaxies, using particle data.

[*]Associate Research fellow at Department of Statistics, University of Warwick

[†]Lecturer of Statistics at Department of Mathematics, University of Leicester

Keywords and phrases: Bayes Factors, Bayesian Nonparametrics, Bayesian P-Values, Hypothesis Testing, Markov Chain Monte Carlo

1. Introduction. One of the burning questions in science today is the understanding of dark matter. The quantification of the distribution of dark matter in our Universe, at different scales, is of major interest in Cosmology (de Blok, Bosma and McGaugh, 2003; Hayashi, Navarro and Springel, 2007; Roberts and Whitehurst, 1975; Salucci and Burkert, 2000; Sofue and Rubin, 2001). At scales of individual galaxies, the estimation of the density of the gravitational mass of luminous as well as dark matter content of these systems, is the relevant version of this exercise. Readily available data on galactic images, i.e. photometric observations from galaxies, can in principle, be astronomically modelled to quantify the gravitational mass density of the luminous matter in the galaxy, (Bell and de Jong, 2001; Gallazzi and Bell, 2009); such luminous matter is however, only a minor fraction of the total that is responsible for the gravitational field of the galaxy since the major fraction of the galactic gravitational mass is contributed to by dark matter. Thus, the gravitational mass density of luminous matter, along with constraints on the gravitational mass density of the dark matter content, if available, can help the learning of the total gravitational mass density. However, the learning of this physical density is difficult in light of the sparse and noisy measurements that are available, coercing the practitioner to resort to undertaking simplifying model assumptions. In this paper, we present a new way of quantifying the relative support to such a model assumption in two independent data sets, by comparing the posterior probabilities of models given the two data sets.

Model selection is a very common exercise faced by practitioners of different disciplines, and substantial literature exists in this field (Barbieri and Berger, 2004; Berger and Pericchi, 2001; Casella et al., 2009a; Chipman, George and McCulloch, 2001; Ghosh and Samanta, 2001; Kass and Raftery, 1995; O'Hagan, 1995). In this context, some advantages of Bayesian approaches, over frequentist methods has been reported (Berger and Pericchi, 2004; Robert, 2001).

Much has been discussed in the literature to deal with the computational challenge of Bayes factors (Casella et al., 2009b; Chib and Jeliazkov, 2001; Han and Carlin, 2000, to name a few). At the same time, methods have been advanced as possible resolutions when faced with the challenge of improper priors on the system variables (Aitkin, 1991; Berger and Pericchi, 1996a; O'Hagan, 1995). However, the computation of posterior, intrinsic or fractional Bayes factors persist as a challenge, especially in the context of discrete, non-parametric and multimodal inference on a high-dimensional state space (Link and Barker, 2006).

This paper demonstrates a novel methodology for quantifying support in two (or more) independent data sets from the same galaxy, for the hypothesis that the system state space admits a certain symmetry - namely, isotropy. Assuming isotropy implies that the probability density function in this high-dimensional state space of the system, depends only on the magnitude of the state space vector, and does not depend on the inclination of this vector to a chosen direction.

The null is nested within the alternative in this problem. At the same time, one of the data sets is small and the other large. Also, the dimensionalities of the model parameter vectors sought under the different hypotheses, are also different. Indeed in such a case, testing with multiple hypotheses can be performed, such that one hypothesis suggests that the state space that one of the data sets is sampled from, admits this symmetry while the other hypothesis suggests the same for the other data set (considering the case of 2 available data sets). However, in the two cases, little prior information are available on the system parameter vectors. The priors on the model parameters are in fact non-informative (uniform), marked by unknown multiplicative constants. Then, as discussed in (Berger and Pericchi, 1996b; Kass and Raftery, 1995), this results in the ratio of the predictive densities of the data sets, under the models, being defined only up to the ratio of these unknown multiplicative constants. Thus, the Bayes factor is rendered arbi-

trary. The computation of the Bayes factors is then, in principle possible with posterior Bayes factors (Aitkin, 1991), intrinsic Bayes factors (Berger and Pericchi, 1996a,b) or with fractional Bayes factors (O'Hagan, 1995).

However in this real-life problem that we discuss here, the implementation of Bayes factors gets difficult given that the relative support to an undertaken assumption in two distinct data sets is sought by comparing the posterior probability of models that are characterised by different numbers of parameters. Secondly, the more complex model in which the simpler model (the null) is nested in this situation, is intractable. The null or the simpler model assumes the model parameter space to bear a simple symmetry, thus rendering posterior computation possible. On the contrary, the alternative (or the more complex model) does not constrain the geometry of the model parameter space in any way. Thus, the simpler model is nested within the more complex model. However, under the alternative hypothesis, i.e. in lieu of such simplifying assumptions, it is not possible to compute the posterior probability of the model parameter vectors. This situation is in principle caused by the complexity of the system, compounded by the paucity of measurements. In the test presented here, learning can be performed - relatively easily under the null - and thereafter, support in the data for the assumption about the symmetry of the model parameter space is quantified.

Lastly, it is acutely challenging in this high-dimensional non-parametric situation, to achieve intrinsic priors (Berger and Pericchi, 1996a) with imaginary training data sets. Such training data can be generated from the posterior predictive distribution under the null (Perez and Berger, 2002), and is subsequently used to train the prior under the more complex model, to implement it in the computation of the marginal density of the real data. However, the computational difficulty involved in the training of the intrinsic prior even under the null, in this high-dimensional setup, is discouragingly daunting.

The implementation of real training data is not possible either, given that any method implemented to learn the gravitational mass density - particularly the similarly motivated Bayesian methodology discussed below - might be misled by undersampled data that is likely to result if subsamples are taken from the already small observed data sets that typify this application.

It was presented in the last paragraph that the generated training data cannot be of practical use in real-life applications of the kind we consider here. However, the implementation of real training data, for the purposes of achieving intrinsic priors, is not possible either. Such real training data is obtained as a subsample of the available data set. Given that for the current application, the data sets are typically small to begin with, estimation of the unknown model parameter vectors using a still smaller (sub)sample of measurements might be risky in terms of convergence of undertaken inference methods.

The difficulties with the computation of Bayes factors that we have delineated above, motivates the need for a test that allows for computation of the comparative support in an available data set for one hypothesis to another, and is operational even with non-informative priors, without needing to invoke any other than the available data set, in the computation of the posterior probability distribution of the hypothesis given the data. The test is also motivated to work irrespective of the dimensionality of parameter spaces. Motivated by this framework, this paper introduces a new nonparametric test of hypothesis that tests for the existence of global symmetries of the phase space that the available data are sampled from. The test works in parameter space, in the context of non-parametric Bayesian inference when for the two or more cases of differently sized samples, little and/or differential prior information are available.

This new test involves partitioning the space of the model parameter vectors such that one of the partitions - the subspace T - contains those parameters for which the posterior probability density given

the data, exceeds the maximal posterior density that can be achieved under the null. Outside T lie the parameter vectors for which the posterior probability falls below this maximal posterior under the null. The model parameter vectors that lie in T are those that underlie the support in the data against the null. Thus, the probability of the null given a data set is the complement of the posterior probability density integrated over the subspace T (instead of over the whole of the parameter space). Such an integral can also be viewed as the probability that a model parameter vector is in the subspace T. In this treatment, the probability of a null given the data that is computationally simple to achieve, in the context of Bayesian nonparametric inference in high-dimensional state spaces.

The paper is organised as follows. Section 2 discusses the application, in the context of which the new test is introduced. The formulation of a phase space pdf as an isotropic scalar-valued function of the velocity and location vectors of a galactic particle, is discussed in Section 3.1. The details of the Bayesian estimation of the unknown parameters of the galaxy, is discussed in Section 3.2. The estimated functions of a real galaxy are presented in Section 3.3. The null hypotheses that we test are motivated in Section 4. In Section 4.1, we discuss the availability of priors on the unknown functions in the relevant literature and how this affects the testing. Thereafter, shortcomings of the Bayes factor computations in the context of this application are delineated in Section 4.2. In Section 5.2, the outline of the new test is presented; in particular, Section 5.4 is devoted to a detailed discussion about the implementation of this new methodology. The test is illustrated on simulated data in Section 6 and on real data in Section 7. The paper is concluded with a discourse on the implications of the results, in Section 8.

2. The application at hand. As discussed above, it is difficult to learn the gravitational mass density $\rho(\mathbf{x})$ of dark+luminous matter in galaxies, where $\rho(\mathbf{x}) > 0$, $\mathbf{X} \in \mathbb{R}^3$. $\rho(\mathbf{x})$ is positive definite, implying that in any infinitesimally small volume $d\mathbf{x}$ inside the galaxy, gravitational mass is non-zero (and positive). Other physically motivated constraints on $\rho(\mathbf{x})$ will be discussed in the next section in the context of learning this function, given the data at hand. While photometric measurements are more readily available, direct detection of dark matter has hitherto been impossible, implying that measurements that can allow for quantification of the gravitational mass density of dark matter, are not achievable. Instead, there are effects of the total (dark+luminous) gravitational mass that can be measured, though astronomical measurements that bear signature of such effects are hard to achieve in "early-type" galaxies, the observed image of which is typically elliptical in shape[1]. Of some such astronomical measurements, noisy and partially missing velocities of individual galactic particles have been implemented to learn $\rho(\mathbf{x})$ (Chakrabarty and Raychaudhury, 2008; Côté et al., 2001; Genzel et al., 2003). From this, when the astronomically modelled gravitational mass density of luminous matter, $\rho_L(\mathbf{x})$, is subtracted, the density of the gravitational mass of the dark matter content of early-type galaxies can be learnt. In this paradigm of learning $\rho(\mathbf{x})$, the data is referred to as partially missing since the noisy measurements of only one component, namely, the component along the line-of-sight that joins the observer to the particle - of the three-dimensional velocity vector of galactic particles, are typically available. We view these measurables as sampled from the phase space $\mathcal{W}$ of the system, where $\mathcal{W}$ is the space of all the states that the system can achieve. For a galaxy, $\mathcal{W}$ is the space of the spatial vectors and velocity vectors of all galactic particles. On other occasions, the dispersion of the line-of-sight component of the velocity vector comprises the measurement. In either case, such kinematic data is expected to track the gravitational field of the galaxy in which the sampled particles play.

[1]The intrinsic global morphology of such "early-type" galaxies is approximated as a triaxial ellipsoid; in this paper we discuss gravitational mass density determination of such galaxies only.

We ask the question that if the available data include velocity measurements of particles in distinct samples that live in mutually disjoint volumes of $\mathcal{W}$, what are the implications for the gravitational mass density estimated using such data sets? One major implication is that the estimate of $\rho(\mathbf{x})$ obtained using one data set will be in general, different from that obtained using another data set. Of course, for the same system, distinct $\rho(\mathbf{x})$ are impossible - in fact, such a fallacious result, if achieved, can be explained as due to the fact that the used data sets have been drawn from mutually insulated volumes of the galactic phase space, the pdfs of which do not concur. Such is possible, if the galactic phase space is characterised by disjoint volumes, motions in which do not communicate with each other. This is a viable scenario in non-linear dynamics even for systems with moderate complexity; it is possible that the two types of galactic particles, the data of which are measured, live in distinct basins of attraction that characterise the galactic phase space (Thompson and Stewart, 2001).

In particular, we configure the question of unequal estimates of gravitational mass density functions, using the available data sets, to the context of the real galaxy NGC 3379. For this system, multiple data sets are measured for two distinct types of galactic particles (Bergond et al., 2006; Douglas et al., 2007).

3. Estimating gravitational mass density. The 3-D spatial location vector of a particle resident in a galaxy is written as $\mathbf{X} = (X_1, X_2, X_3)^T$, where only the coordinates of the image of the particle - X_1 and X_2 - can be measured. Also, the 3-D velocity vector of a particle is $\mathbf{V} = (V_1, V_2, V_3)^T$, with only the component V_3 being a measurable, i.e. we can only measure the speed with which the particle is approaching us (the observer) or receding from us. The galactic phase space $\mathcal{W}$ is the space of $\mathbf{V}$ and $\mathbf{X}$ of all galactic particles. Thus, $\mathcal{W} \subseteq \mathbb{R}^6$.

3.1. *Assuming an isotropic phase space pdf.* If we are convinced that motion tracks the gravitational field due to a given gravitational mass density function, then we can write down $\mathbf{v} = \xi[\rho(\mathbf{x})]$, where $\xi(\cdot)$ is a function of $\rho(\mathbf{x})$. In Chakrabarty and Portegies Zwart (2004); D.Chakrabarty (2009), the effort has been to learn $\rho(\mathbf{x})$ from the data using non-parametric modelling. The aim in these work is to make inference on the gravitational mass density $\rho(\mathbf{x})$ given observations of only 1 component V_3 out of 3 components of the velocity vector of particles resident in the system, where the dependence of V_3 on $\rho(\mathbf{x})$ is intuitively motivated.

In order to render the problem tractable, we need to undertake some simplifying model assumptions. It is useful in such situations to assume something simple about symmetries of the system state space. The simplest such assumption considers phase space $\mathcal{W}$ to be isotropic, i.e. for the phase space probability density function $f(\mathbf{x}, \mathbf{v}) \geq 0$, to be isotropic in $\mathbf{X}$ and $\mathbf{V}$. Thus, the phase space pdf is $f : \mathcal{W} \longrightarrow \mathbb{R}_{\geq 0}$. In general, the pdf of $\mathcal{W}$ has a time dependence, but here we assume the system to have attained stationarity, i.e. for the phase space pdf to be independent of time. Thus, one model assumption that we undertake is that of stationarity of the phase space pdf. Such is typically undertaken in the modelling of galaxies since deviation from stationarity cannot be checked in the available data since the data sets offer snapshots at a (nearly) unique time point in the evolution of the system, the time scale of which $\sim 10^9$ years, (Binney and Tremaine, 1987). There are reports of observational signatures of non-stationarity; in NGC 3379 these could be considered to include the X-ray image of NGC 3379 (Pellegrini and Ciotti, 2006)[2]. All in all, at the level of this paper, we cannot say that deviation from equilibrium is ruled out.

[2] Pellegrini and Ciotti (2006) report the gravitational mass of the galaxy modelled using X-ray measurements, as overestimated by a factor of about 2, compared to hydrodynamical modelling that assume hydrostatic equilibrium. Pellegrini and Ciotti (2006) invoke the imaged out-flowing nature of the hot gas from this galaxy to support deviation from hydrostatic equilibrium. However, the contribution of such deviations from hydrostatic equilibrium, towards the estimate of gravitational mass has not

Thus we state that it is under this model assumption of stationarity, that we interpret all results. The phase space *pdf* is then a scalar-valued, function of 2 vectors, $\mathbf{X}$ and $\mathbf{V}$.

Now, if a general real-valued function $g(\cdot,\cdot)$ of two vectors $\mathbf{a},\mathbf{b}\in\mathbb{R}^m$, is an isotropic function, then $g(\mathbf{a},\mathbf{b})=g(\mathbf{Qa},\mathbf{Qb})$, for any orthogonal transformation matrix $\mathbf{Q}$ (Truesdell, Noll and Antman, 2004; Wang, 1969). We recall from the theory of scalar valued functions of two vectors, that if $g(\cdot,\cdot)$ is an isotropic function, its set of invariants (with respect to $\mathbf{Q}$) is $\Upsilon_Q=\{\mathbf{a}\cdot\mathbf{a},\mathbf{b}\cdot\mathbf{b},\mathbf{a}\cdot\mathbf{b}\}$. Then, the isotropic function of two vectors, $g(\mathbf{a},\mathbf{b})$, admits the representation $g(\Upsilon_Q)$ (Liu, 2002; Truesdell, Noll and Antman, 2004). In general, Υ_Q is the sum of a function of $\mathbf{a}\cdot\mathbf{a}$ with a function of $\mathbf{b}\cdot\mathbf{b}$ and a function of $\mathbf{a}\cdot\mathbf{b}$. If $\mathbf{a}$ and $\mathbf{b}$ are orthogonal vectors, $\mathbf{a}\cdot\mathbf{b}$=0.

Then, it follows that

- under the assumption that the phase space *pdf* is an isotropic scalar valued function of 2 vectors, $\mathbf{x}$ and $\mathbf{v}$, it admits the representation $f(\epsilon)$,
- if we can write a general phase space *pdf* as $f(\epsilon)$, it implies that it is an isotropic scalar valued function of $\mathbf{x}$ and $\mathbf{v}$,

where ϵ is a sum of a function of $\mathbf{x}\cdot\mathbf{x}$ with a function of $\mathbf{v}\cdot\mathbf{v}$. (The inner product $\mathbf{X}\cdot\mathbf{V}$=0),i.e.

$$\epsilon=\Phi(\sqrt{\mathbf{x}\cdot\mathbf{x}})+\eta(\mathbf{v}\cdot\mathbf{v}), \tag{3.1}$$

with $\Phi(\cdot)$ and $\eta(\cdot)$ identified as some real-valued functions.

One physical interpretation of the above representation is that the phase space *pdf* is isotropic if and only if it can be written as a function of the value ϵ of the particle energy, where energy of a particle with location $\mathbf{X}=\mathbf{x}$ and velocity $\mathbf{V}=\mathbf{v}$ is the function $E(x^2,v^2)$, i.e.

$$\epsilon=E(x^2,v^2). \tag{3.2}$$

N.B. the symbol E is to be distinguished from expectation. Now total energy is the sum of potential and kinetic energies. Thus, to allow for the physical interpretation of ϵ as the value of the particle energy, we choose $\Phi(\sqrt{\mathbf{x}\cdot\mathbf{x}})$ to be the gravitational potential energy, (written as a function of $\mathbf{x}\cdot\mathbf{x}=x_1^2+x_2^2+x_3^2$) and the kinetic energy $\eta(\mathbf{v}\cdot\mathbf{v})=\mathbf{v}\cdot\mathbf{v}/2$[3]. It merits mention that $\rho(\mathbf{x})$ is a known function of $\Phi(\mathbf{x})$ as given by a standard equation of Physics - the Poisson equation:

$$\nabla^2\Phi(\mathbf{x})=-4\pi G\rho(\mathbf{x}), \tag{3.3}$$

where ∇^2 is the Laplacian operator and G is a known constant.

However, in lieu of information about the nature of the phase space density, and given at least the moderate level of complexity that is envisaged to manifest in phase spaces of galaxies, there is no justification for opting for an isotropic, over an anisotropic prescription for galactic phase spaces. There is a strong logistical motivation for doing so however - the assumption of an isotropic $f(\mathbf{x},\mathbf{v})$ renders the

been satisfyingly addressed.

[3]The above identification of the dependence of f on particle energy is not motivated merely to invoke familiar ideas of physics into the model but is reinforced by the fact that phase space *pdf* depends on $\mathbf{X}$ and $\mathbf{V}$ only via constants of motion since $\frac{df}{dt}=0$, i.e. $f(\mathbf{x},\mathbf{v})$ obeys the Collisionless Boltzmann Equation(Binney and Tremaine, 1987). The only constant of motion I that depends on $\mathbf{X}$ and $\mathbf{V}$ - as in $I(\mathbf{x}\cdot\mathbf{x},\mathbf{v}\cdot\mathbf{v})$ - is the particle energy. Thus, an isotropic phase space *pdf* depends on $\epsilon=E(x^2,v^2)$ or any regular function of ϵ. Without loss of generality, for our purpose, we suggest that the isotropic phase space *pdf* is $f(\epsilon)$.

calculations involved in the learning of $\rho(\mathbf{x})$ relatively easier. The details of this learning, given the data and the assumed prescription of isotropy for $f(\mathbf{x}, \mathbf{v})$, are discussed in Section 3.2. In principle, learning $\rho(\mathbf{x})$ after relaxing the assumption of isotropy, under a particular restrictive prescription for including anisotropy, can still be possible with a much harder computational procedure (Chakrabarty and Saha, under preparation). However, it is impossible to learn $\rho(\mathbf{x})$ in the general anisotropic cases, given the typically partially missing nature of the kinematic data, under an unrestricted model for the anisotropy of phase space. Thus, the estimate of $\rho(\mathbf{x})$ is sensitive to the prescription chosen for the phase space symmetry - isotropic or not, and if anisotropic, to the value of the "anisotropy parameter". This problem refers to the lack of identifiability among solutions for $\rho(\mathbf{x})$, caused by lack of knowledge about the state of isotropy of galactic phase spaces. and is referred to as the mass-anisotropy degeneracy in the astrophysical literature (Côté et al., 2001, 2003; Koopmans, 2006; Łokas and Mamon, 2003).

The problem has been a pressing concern of Dekel et al. (2005) when they advance the possibility that anisotropy might be introduced into $f(\mathbf{x}, \mathbf{v})$ of galaxy NGC 3379 due to evolutionary reasons, which Romanowsky et al. (2003) had ignored. Douglas et al. (2007) disagree that this could explain the identification of the low dark matter content in NGC 3379 by Romanowsky et al. (2003) given that the existence of anisotropy was included in the data analysis employed in the earlier work. They stress that "anisotropy parameter" recovered from their kinematic data and that obtained from the simulations presented in Dekel et al. (2005), are similar in nature. [4]. Here, the anisotropy parameter is a parametrisation of the deviation from an isotropic phase space density, defined as $1 - \sigma_k^2(\mathbf{x})/\sigma_3^2(\mathbf{x})$ (Binney and Tremaine, 1987), where $\sigma_k^2(\mathbf{x})$ is the variance in the k-th component of $\mathbf{V}$, $k = 1, 2$.

However, the suggestion advanced in this paper is more general. Firstly, the learning of $\rho(\mathbf{x})$ is non-parametric, so an explicit parametrisation of anisotropy is not of relevance. Importantly, if the test of hypothesis that we develop here, suggests less support in one data set over another, to the model assumption of an isotropic phase space pdf, it would imply that the two data sets have been sampled from distinct phase space density functions that will then be interpreted as fundamentally different. Such unequal pdfs cannot in general be reconciled with each other, merely by adjusting one parameter, the physical interpretation of which is the anisotropy parameter as given in astrophysical literature.

3.2. *Non-parametric Bayesian learning of* $\rho(\mathbf{x})$. In general, the phase space pdf will contain parameters $\boldsymbol{\alpha}$, i.e. the phase space pdf is $f(\mathbf{x}, \mathbf{v}; \boldsymbol{\alpha})$. Given data $\mathbf{D}$ of the observable phase space coordinates of a sample of N_{tot} particles, the likelihood function $\mathcal{L}(\boldsymbol{\alpha}|\mathbf{D})$ can be written using the phase space pdf $f(\mathbf{x}, \mathbf{v}; \boldsymbol{\alpha})$ since by definition, $d\Pr(\mathbf{D}^{(k)}|\boldsymbol{\alpha}) = f(\mathbf{x}^{(k)}, \mathbf{v}^{(k)}; \boldsymbol{\alpha}) d^3\mathbf{x} d^3\mathbf{v}$, where the data vector of the k-th particle is $\mathbf{D}^{(k)} = (\mathbf{x}^{(k)}, \mathbf{v}^{(k)})^T$ and the spatial and velocity vectors of this particle are $\mathbf{x}^{(k)}$ and $\mathbf{v}^{(k)}$ respectively, $k = 1, 2, \ldots, N_{tot}$. However, there are two points of worry at this stage.

Firstly, we want to learn $\rho(\mathbf{x})$ but it would appear that the likelihood proposed above does not include this function in its definition. Generally speaking, in such situations, we would need to embed the sought unknown function ($\rho(\mathbf{x})$) in the definition of the phase space pdf by suggesting that $f(\mathbf{x}, \mathbf{v}; \boldsymbol{\alpha}) \equiv f(\Psi(\rho(\mathbf{x}), \mathbf{v}, \boldsymbol{\alpha}_1), \mathbf{x}, \mathbf{v}, \boldsymbol{\alpha}_2)$, where $\boldsymbol{\alpha} := (\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T)^T$. This calls for identification of a function $\Psi(\rho(\cdot), \cdot, \cdot)$ using domain knowledge. The likelihood would then in general be written as

$$\mathcal{L}(\boldsymbol{\alpha}|\mathbf{D}^{(1)}, \ldots, \mathbf{D}^{(N_{tot})}) \propto \prod_{k=1}^{N_{tot}} f(\Psi(\rho(\mathbf{x}^{(k)}), \mathbf{v}^{(k)}, \boldsymbol{\alpha}_1), \mathbf{x}^{(k)}, \mathbf{v}^{(k)}, \boldsymbol{\alpha}_2) \tag{3.4}$$

[4] Weijmans and et al. (2009) cannot infer the distribution of the total gravitational mass distribution in this galaxy using a parametric methodology since they cannot obtain information on certain model parameters. Coccato et al. (2009) and Pierce and et al. (2006) report the characterisation of this galaxy using available kinematic data.

where the data are considered to be iid.

In our case, we have already identified that the admissible representation of an isotropic phase space pdf, is as a function of particle energy, (Section 3.1), which is itself the sum of potential and kinetic energies. We have defined $\epsilon := E(\mathbf{x}\cdot\mathbf{x}, \mathbf{v}\cdot\mathbf{v}) = \Phi(\sqrt{\mathbf{x}\cdot\mathbf{x}}) + \mathbf{v}\cdot\mathbf{v}/2$, (see Equations 3.1, Equation 3.2), where $\Phi(\mathbf{x})$ depends on the sought unknown function $\rho(\mathbf{x})$ via Poisson equation (see Equation 3.3). Thus, a function that has $\rho(\sqrt{\mathbf{x}\cdot\mathbf{x}})$ embedded in it, is the energy function $E(\mathbf{x}\cdot\mathbf{x}, \mathbf{v}\cdot\mathbf{v})$ that we now re-write as $E(\rho(\sqrt{\mathbf{x}\cdot\mathbf{x}}), \mathbf{v}\cdot\mathbf{v})$, so that

$$\epsilon = E(x^2, v^2) = E(\mathbf{x}\cdot\mathbf{x}, \mathbf{v}\cdot\mathbf{v}) \equiv E(\rho(\sqrt{\mathbf{x}\cdot\mathbf{x}}), \mathbf{v}\cdot\mathbf{v}) \equiv E(\rho(\parallel \mathbf{x} \parallel), \parallel \mathbf{v} \parallel^2) \tag{3.5}$$

where $\parallel \cdot \parallel$ denotes the Euclidian norm of a vector. To summarise, the dependence of $f(\cdot)$ on $\rho(\mathbf{x})$ is ensured by suggesting that the phase space pdf depends on value of energy function, ϵ. Given this, we could suggest that the phase space density is $f(E(\rho(\mathbf{x}), \mathbf{v}^2)), \mathbf{x}, \mathbf{v}, \boldsymbol{\alpha}_2)$. However, here we assume that the phase space is isotropic and encode this assumption by writing the phase space pdf as a function of energy alone, with no dependence on any other variables (see Section 3.1). In fact, we assume that phase space pdf is $f(\epsilon)$, i.e. $f(E(\rho(\parallel \mathbf{x} \parallel), \parallel \mathbf{v} \parallel^2))$ (by Equation 3.5).

Now, there are no universally accepted parametric forms of the gravitational mass density of all matter in galaxies, known of in the astronomical literature. This urges a fully non-parametric model for $\rho(\parallel \mathbf{x} \parallel)$. Following D.Chakrabarty (2009), we adopt a fully discretised model such that the range of $\parallel \mathbf{x} \parallel$ is binned, with $\parallel \mathbf{x} \parallel \in [\parallel \mathbf{x}_i \parallel - \parallel \delta\mathbf{x} \parallel, \parallel \mathbf{x}_i \parallel)$ defining the i-th $\mathbf{X}$-bin of width $\parallel \delta\mathbf{x} \parallel$. We define that when $\parallel \mathbf{x} \parallel$ is in the i-th $\mathbf{X}$-bin, $\rho(\parallel \mathbf{x} \parallel) := \rho_i$, $i = 1, 2, \ldots, N_x$ and define the vector $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_{N_x})^T$. Thus, $\boldsymbol{\rho}$ is the discretised form of the unknown function $\rho(\parallel \mathbf{x} \parallel)$ that we seek to learn from the data. In light of this, we view the energy function as $E(\boldsymbol{\rho}, \parallel \mathbf{v} \parallel^2) = \epsilon$.

In fact, there is again no guideline in the literature as to the nature of dependence of the phase space pdf on energy, motivating a non-parametric model for $f(\epsilon)$ too. The relevant range of energy values is discretised such that $\epsilon \in [\epsilon_j - \delta_\epsilon, \epsilon_j)$ defines the j-th E-bin of width δ_ϵ, inside which, $f(\epsilon) = f_j$, $j = 1, \ldots, N_E$. Thus, we set $f(\epsilon) = f_j$ for ϵ in the j-th E-bin, $j = 1, \ldots, N_E$. We define the vector $\mathbf{f} := (f_1, f_2, \ldots, f_{N_E})^T$. Indeed $\mathbf{f}$ is unknown and we attempt to learn it from the data.

Our second worry is with the suggestion for the likelihood in Equation 3.6. We recall that the data $\mathbf{D}$ is partially missing as it does not comprise measurements of all 6 phase space coordinates; in fact $\mathbf{D} := \{x_1^{(k)}, x_2^{(k)}, v_3^{(k)}\}_{k=1}^{N_{tot}}$. Thus, in addition to the line-of-sight speed V_3, the measurables include the spatial locations X_1, X_2 of the particle on the image of the galaxy. Let $(X_1, X_2, V_3)^T \in \mathcal{M} \subset \mathcal{W}$. The data is then sampled from the pdf $\nu(x_1, x_2, v_3)$ of the subspace $\mathcal{M} \subset \mathcal{W}$, so that likelihood should be defined in terms of $\nu(x_1, x_2, v_3)$. In other words, the likelihood function cannot contain the unobservable variables X_3, V_1, V_2. Therefore, the definition of likelihood in Equation 3.4 has to be modified and written in terms of the projected phase space pdf $\nu(x_1^{(k)}, x_2^{(k)}, v_3^{(k)})$, where this projected pdf results from the projection of $f(E)$ onto the space of the observables, $\mathcal{M}$. Assuming the observed vectors $(x_1^{(k)}, x_2^{(k)}, v_3^{(k)})^T$, $k = 1, \ldots, N_{tot}$, to be $i.i.d$, the likelihood function is then defined as

$$\mathcal{L}(\rho_1, \ldots, \rho_{N_x}, f_1, \ldots, f_{N_E} | \mathbf{D}) = \prod_{k=1}^{N_{tot}} \nu(x_1^{(k)}, x_2^{(k)}, v_3^{(k)}), \tag{3.6}$$

where the projected phase space pdf for the k-th data point, $\nu(x_1^{(k)}, x_2^{(k)}, v_3^{(k)})$, results from projecting

the phase space *pdf* $f(\epsilon)$ onto the subspace $\mathcal{M}$, as in

$$\begin{aligned} \nu(x_1^{(k)}, x_2^{(k)}, v_3^{(k)}) &= \int_{X_3}\int_{V_1}\int_{V_2} f(\epsilon) dx_3 dv_1 dv_2, \\ \text{i.e.} \quad \nu(x_1^{(k)}, x_2^{(k)}, v_3^{(k)}) &= \int_{X_3}\int_{V_1}\int_{V_2} f(E(\rho(x_1^{(k)}, x_2^{(k)}, x_3), v_1, v_2, v_3^{(k)}) dx_3 dv_1 dv_2, \end{aligned} \tag{3.7}$$

where this projected phase space *pdf* is normalised at each iteration by a constant given by integrating $\nu(\cdot,\cdot,\cdot)$ over all possible values of the observables V_1, V_2, X_3. It merits mention that E inside the integral is the energy function, not to be confused with expectation. In fact,

$$\begin{aligned} \epsilon &:= E(\rho(x_1^{(k)}, x_2^{(k)}, x_3), v_1, v_2, v_3^{(k)}) \\ &= \Phi\left(\sqrt{\{x_1^{(k)}\}^2 + \{x_2^{(k)}\}^2 + \{x_3\}^2}\right) \Big| \rho\left(\sqrt{\{x_1^{(k)}\}^2 + \{x_2^{(k)}\}^2 + \{x_3\}^2}\right) + \\ &\quad \frac{(\{v_1\}^2 + \{v_2\}^2 + \{v_3^{(k)}\}^2)}{2} \end{aligned} \tag{3.8}$$

In the second step of Equation 3.8, elaboration of energy as sum of the potential and kinetic energies invokes the idea that the potential $\Phi(\cdot)$ is determined via Poisson equation (Equation 3.3), given the gravitational mass density function $\rho(\cdot)$.

The projection from $\mathcal{W}$ onto $\mathcal{M}$ in Equation 3.7 is performed by decomposing the integrals on the RHS as a sum of the contributions to the integrals, from individual E-bins. The mapping from the space $\mathcal{W}/\mathcal{M}$ to the range of ϵ in the j-th E-bin, is identified for any k (see Chakrabarty and Portegies Zwart, 2004, for details), $j = 1, 2, \ldots, N_{eng}$. Summing over all j's then provides $\nu(x_1^{(k)}, x_2^{(k)}, v_3^{(k)})$.

Owing to lack of consensus in the literature on the nature of the gravitational mass densities in galaxies and their phase space *pdf*s, we choose to impose uniform priors for ρ_h and f_j, $h = 1, \ldots, N_x$, $j = 1, \ldots, N_{eng}$. These are respectively, $\pi_0(\rho_h) = \mathcal{U}[\rho_{lo}^{(h)}, \rho_{hi}^{(h)}]$ and $\pi_0(f_j) = \mathcal{U}[0,1]$, . Here $\rho_{lo}^{(h)}, \rho_{hi}^{(h)}$ are experimentally chosen constants and the prior on the phase space density is uniform in [0,1] since $f_{N_{eng}}$ is normalised to be 1 for the most bound orbit, i.e. for the maximum value of ϵ[5].

The priors are used along with the likelihood function (defined above in Equation 3.6) in Bayes rule to write the posterior probability density

$$\pi_1(\rho_1, \ldots, \rho_{N_x}, f_1, \ldots, f_{N_E}|\mathbf{D}) \propto \mathcal{L}(\rho_1, \ldots, \rho_{N_x}, f_1, \ldots, f_{N_E}|\mathbf{D}) \times \prod_{h=1}^{N_x} \pi_0(\rho_h) \prod_{j=1}^{N_E} \pi_0(f_j).$$

Above, the posterior density is referred to as $\pi_1(\boldsymbol{\rho}, \mathbf{f}|\mathbf{D})$ to distinguish it from the posterior probability $\pi(\boldsymbol{\rho}, \mathbf{f}|\mathbf{D})$ that we actually sample from, after convolving $\pi_1(\cdot,\cdot|\cdot)$ with the distribution of error in the measurements of V_3, $f_{error}(v_3)$. The measurement errors in X_1 and X_2 are stated by astronomers to be small enough to be ignored. Inputs from astronomers involved in the observed data set are considered when modelling the error distribution. Typically, the error distribution is considered Gaussian with zero

[5] It will be seen in the presentation of our results that ϵ is presented as normalised by its maximum possible value $\Phi_0 := \Phi(0)$. $\Phi(0)$ is determined by $\Phi(r)$ which in turn is determined by using the learnt $\boldsymbol{\rho}$ in Poisson equation (Equation 3.3). Thus, the normalised value energy $\epsilon \in [0,1]$.

mean and variance that is suggested by the relevant observational astronomer(s). Thus,

$$\pi(\boldsymbol{\rho}, \mathbf{f}|\mathbf{D}) = \pi_1(\boldsymbol{\rho}, \mathbf{f}|\mathbf{D}) * f_{error}(v_3).$$

Then $\mathbf{f}$ and $\boldsymbol{\rho}$ are learnt by sampling from this posterior, using adaptive random-walk Metropolis-Hastings (Haario et al., 2006). Some of the approximations that underlie sampling from this posterior are now discussed, followed by a brief discussion of the inference.

This is the approach used by Chakrabarty (2006); Chakrabarty and Portegies Zwart (2004) and Chakrabarty and Raychaudhury (2008) who assumed an isotropic phase space pdf. Now, the undertaken model assumption of phase space isotropy also allows for the identification of $\sqrt{\mathbf{xx}} = \| \mathbf{x} \|$, with the spherical radius r, where $r^{(k)} \equiv \sqrt{\{x_1^{(k)}\}^2 + \{x_2^{(k)}\}^2 + \{x_3\}^2}$, the radial location of the k-th particle from the system centre. In other words, the assumption of an isotropic phase space is inclusive of a spherical spatial geometry. Then identifying $\Phi(\cdot)$ with the gravitational potential energy , the connection between $\Phi(r)$ and $\rho(r)$ (Poisson equation) is recalled as $\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\Phi(r)}{dr}\right) = -4\pi G\rho(r)$ in this geometry. Then the h-th $\mathbf{X}$-bin discussed above is synonymous to the h-th radial bin, $h = 1, \ldots, N_x$.

The unknown functions abide by the physically motivated constraints that $\rho(r) \geq 0$, $f(\epsilon) \geq 0$ and $\frac{d\rho(r)}{dr} \leq 0$. The last constraint is intuitively motivated as valid in a gravitationally bound spherical system that we model the galaxy to be. We view the galaxy as being built by stacking spherical layers on top of each other. Then owing to gravity being a force that attracts mass towards the centre, the compactness in the packaging of mass is higher near the centre than away from the centre. In other words, $\rho(r)$ increases as r decreases, in general (Binney and Tremaine, 1987). Following the trend of the various phase space pdfs discussed in astronomical literature (Binney and Tremaine, 1987), $f(E)$ is also treated as a monotonically increasing function of energy E in this implementation of the methodology.

In the adopted discrete model, a simultaneous learning of discretised versions of two univariate functions $f(\epsilon)$ and $\rho(r)$ is attempted, i.e. the target is to learn the vector $\boldsymbol{\rho} := (\rho_1, \ldots, \rho_{N_x})^T$ and the vector $\mathbf{f} := (f_1, \ldots, f_{N_E})^T$ as proxies for the unknown functions. Given that the gravitational mass density is non-negative and monotonically non-increasing function of r, $\boldsymbol{\rho} \in \mathcal{R} \subset \mathbb{R}^{N_x}$. Again, given that $f(\epsilon) \geq 0$, $\mathbf{f} \in \mathcal{F} \subset \mathbb{R}^{N_E}$.

In our implementation, to ensure monotonic decline in the gravitational mass density function, it is the difference $\Delta_\rho^{(h)}$ between the gravitational mass densities in the h-th and $h+1$-th radial bins that is proposed, as $\tilde{\Delta}_\rho^{(h)} \sim \mathcal{N}_F(\Delta_\rho^{(h)}, s^2)$, where $\mathcal{N}_F(c, d)$ is the folded normal distribution (Leone, Nottingham and Nelson, 1961) with mean c and variance d, $c \geq 0$, $d > 0$. The motivation behind this choice of the proposal density is that over the support $\mathbb{R}_{\geq 0}$, it is a relatively easy density to sample from, while satisfying the requirement that in general, $\Pr(\tilde{\Delta}_\rho^{(h)}) > 0$ when $\tilde{\Delta}_\rho^{(h)} = 0$. Here, the current difference between the gravitational mass density values in the h and $h+1$-th radial bins is $\Delta_\rho^{(h)} := \rho_h - \rho_{h+1}$. The variance s^2 is the empirical variance of $\Delta_\rho^{(h)}$, computed using values of this difference variable from step number t_0 to $t-1$, where t is the current step number and t_0 is chosen experimentally to be post-burnin (Haario et al., 2006), $h = 1, \ldots, N_x$. Here ρ_{N_x+1} and $\tilde{\rho}_{N_x+1}$ are defined as 0. Then as h is varied from N_x to 1, the proposed h-th component of the unknown gravitational mass density vector is $\tilde{\rho}_h = \tilde{\rho}_{h+1} + \tilde{\Delta}_\rho^{(h)}$. f_j is updated similarly, while following an imposed constraint that $f(E)$ is monotonically non-decreasing with energy E, $j = 1, \ldots, N_{eng}$. However, unlike in the case of the updating

of the gravitational mass density value, a monotonically non-decreasing phase space pdf with energy is not motivated by any physically justifiable constraints, though astronomical literature often suggests phase space density functions that typically increase with energy (Binney and Tremaine, 1987).

Thus, we realise that when the data comprises even partially missing information on the state space coordinates of a system, this methodology allows for the formulation of the likelihood of the parameters describing the sought system function, in terms of the projection of the pdf of the system state space onto the space of the observables, as long as the unknown function can be embedded within the definition of this pdf. Such an approach is useful in contexts similar to what we work with here, for applications other than the current one. A notable feature of this methodology is that even when the data are missing, i.e. sampled from a sub-space of the system state space, it is possible to write the likelihood function as a product of the values of the sub-space density, at each data point, where the density of this sub-space is obtained by integrating the unobserved variables out of the state space pdf.

3.3. *Bayesian learning of model parameter vectors in real galaxy NGC 3379.* NGC 3379 is one of the few elliptical galaxies, for which kinematic information is available for individual members of two different populations of galactic particles - referred to as planetary nebulae (PNe) and globular clusters (GCs) - over an extensive radial range spanning the outer parts of the galaxy. The data used in the work include measurements of X_1, X_2 and V_3 of 164 PNe reported by Douglas et al. (2007) and of 29 GCs by Bergond et al. (2006)[6]. We refer to the PNe data set as $\mathbf{D}_1$ and the GC data set as $\mathbf{D}_2$, with respective sample sizes of N_1=164 and N_2=29. The learning of the model parameter vectors $\mathbf{f}$ and$\boldsymbol{\rho}$ is performed in a high-dimensional state space (dimensionality = $N_{eng} + N_x$). The traces of the posterior given either data are presented in Figure 1. The marginal posterior densities given either data are shown for one component of the learnt $\boldsymbol{\rho}$ vector, namely rho_6 (Figure 3). The marginal densities are noted to be markedly multimodal.

In Figure 2, we present the vectors $\boldsymbol{\rho}^{(i)} := (\rho_1^{(i)}, \ldots, \rho_{N_x}^{(i)})^T$ and $\mathbf{f}^{(i)} := (f_1^{(i)}, \ldots, f_{N_E}^{(i)})^T$, learnt from using the i-th data $\mathbf{D}_i$, $i = 1, 2$. This estimation is performed using the aforementioned Bayesian nonparametric methodology, under the model assumption of an isotropic phase space pdf, i.e. a phase space density that is expressed as $f(\epsilon)$ and approximated in the discretised model of CHASSIS as the vector $\mathbf{f}$, (the j-th component of which is the value of the phase space density in the j-th energy-bin) and the vector $\boldsymbol{\rho}$, (the h-th component of which is the value of the phase space density in the h-th radial-bin). The learnt 95% HPDs are represented as error bars and the modal values are shown as open circles.

4. Testing for isotropy. The apparently inconsistent $\boldsymbol{\rho}$ learnt from two different kinematic data sets of two different types of galactic particles in a real galaxy, (Figure 2), translates to multiple gravitational potential estimates for the same galaxy, which is of course not physically plausible. Thus, the total gravitational mass in a galaxy can be misidentified, leading to erroneous estimation of the fraction of the galactic gravitational mass that is contributed to by dark matter which if applicable to a sample of galaxies, can in turn bias cosmological ideas about the distribution of dark matter on galactic scales.

Thus, we have identified the risk in the implementation of kinematic data in learning galactic gravitational potential. In appreciation of this serious risk, we need to check
–if the problem is inherent to the very notion of attempting the estimation of gravitational mass density using the available data on phase space coordinates of sampled galactic particles, or

[6]We use the velocities of only 29 of all the GCs that Bergond et al. (2006) advance, as these are identified as unambiguously bound to the gravitational field of NGC 3379, as distinguished by others that might be in the shared by the fields of this galaxy and its neighbours.

TABLE 1

Table displaying seeds that individual chains run with data $\mathbf{D}_1$ and $\mathbf{D}_2$ are started with. The initial choice of the gravitational mass density function is one that is sometimes used in astrophysical literature, $\frac{\rho_0}{\left(\frac{r}{r_c}\right)^{\alpha_1}\left(1+\frac{r}{r_c}\right)^{\alpha_2}}$. The starting phase space density is chosen to be either of the form $\exp^{\epsilon}$ or ϵ^{β}. Here the parameters $\rho_0, \alpha_1, \alpha_2, r_c, \beta \in \mathbb{R}_{\geq 0}$. The chains run with PNe data $\mathbf{D}_1$ are assigned names characterised by the prefix "PNe-RUN", while runs performed with the GC data $\mathbf{D}_2$ are labelled with prefix "GC-RUN".

Name	ρ_o	r_c (kpc)	α_1	α_2	f_{seed}
$PNe-RUN\ I$	10^3	30	2.8	1	$=\exp(\epsilon)$
$PNe-RUN\ II$	10^{10}	20	3.8	2	$=\epsilon^3$
$PNe-RUN\ III$	10^{14}	10	1.8	3	$=\epsilon^5$
$GC-RUN\ I$	10^5	30	2.8	1	$=\exp(\epsilon)$
$GC-RUN\ II$	10^8	5	3.6	2	$=\epsilon^5$
$GC-RUN\ III$	10^{10}	10	3.2	3	$=\epsilon^2$

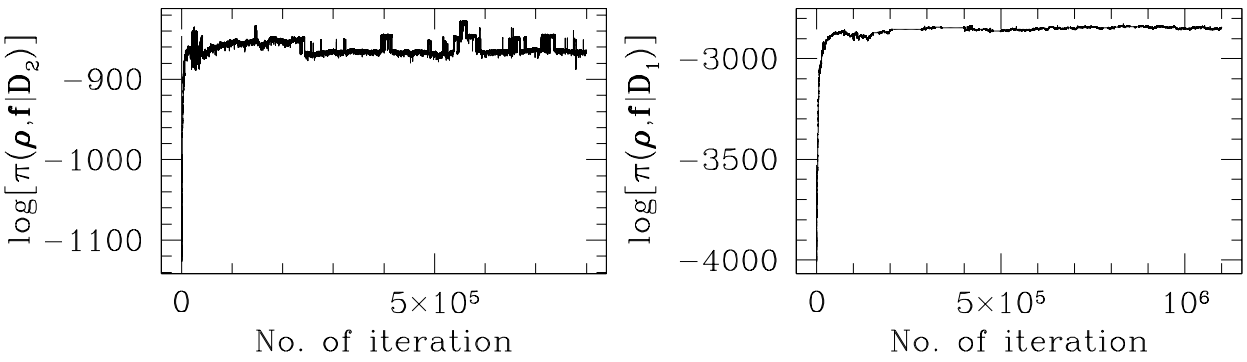

FIG 1. *Logarithm of the joint posterior probability density of model parameter vectors $\mathbf{f}$ and $\boldsymbol{\rho}$ given the two sets of real data - $\pi(\mathbf{f}, \boldsymbol{\rho}|\mathbf{D}_1)$ on the right and $\pi(\mathbf{f}, \boldsymbol{\rho}|\mathbf{D}_1)$ on the left.*

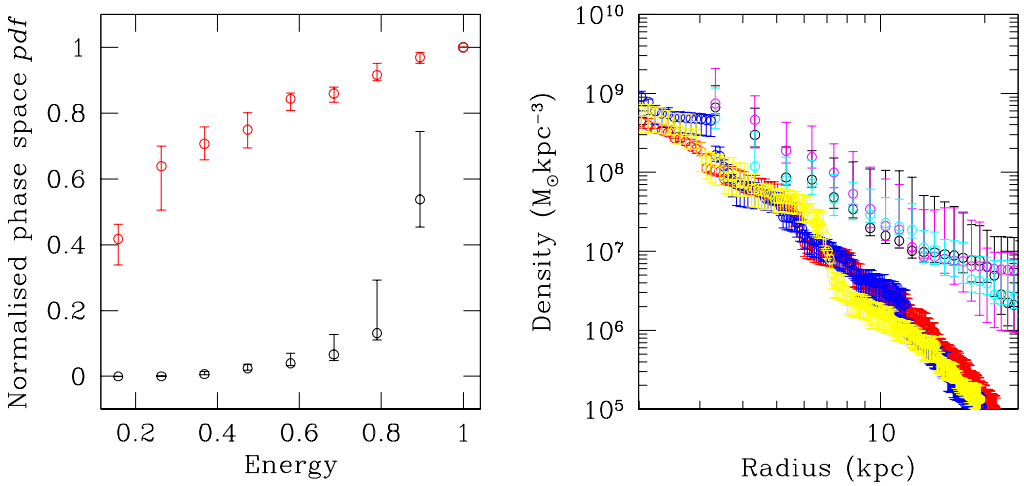

FIG 2. *Right panel: density profiles recovered from the chains run with real data* $\mathbf{D}_1$ *(data on PNe) and* $\mathbf{D}_2$ *(data on GCs). The* $\boldsymbol{\rho}_1$ *learnt from the chains run with PNe data* $\mathbf{D}_1$ *are shown in red, yellow and blue while* $\boldsymbol{\rho}_2$ *learnt with the GC data* $\mathbf{D}_2$ *kinematics are in black, magenta and cyan. As is apparent,* $\boldsymbol{\rho}$ *vectors learnt using distinct data sets are not consistent with each other. Left panel: isotropic (normalised), phase space density vector* $\mathbf{f}_1$ *recovered from* $PNe-RUN\ I$ *(in red) and* $\mathbf{f}_2$ *from* $GC-RUN\ I$ *(in black), plotted in corresponding energy-bins.*

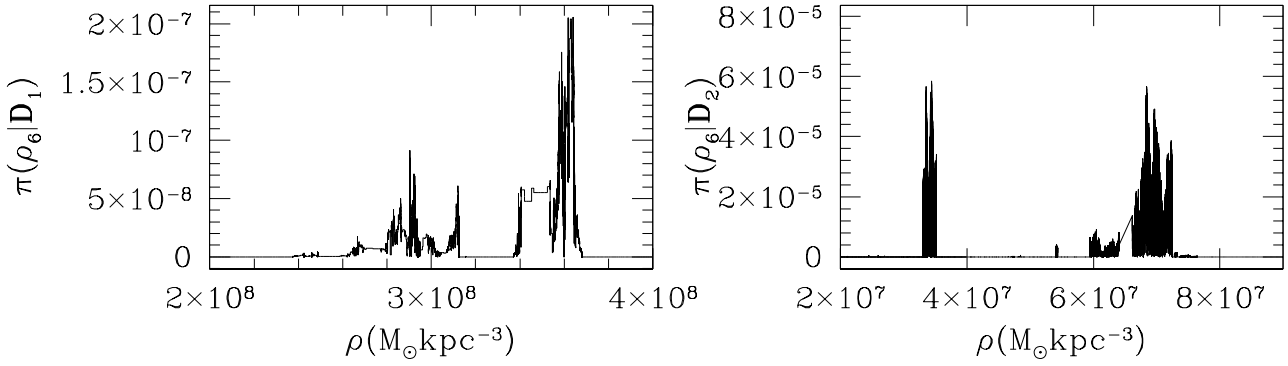

FIG 3. *Marginal posterior probability density of the 6-th component of the* $\boldsymbol{\rho}$ *vector learnt using observed data* $\mathbf{D}_1$ *(left) and* $\mathbf{D}_2$ *(right). The plots manifest multimodality of the marginal posterior densities of the 6-th component of the model parameter vector* $\boldsymbol{\rho}$*.*

–if there is a problem in our Bayesian methodology discussed above, or
–if the data implemented in this example are biased to produce the recovered inconsistency between $\boldsymbol{\rho}^{(1)}$ and $\boldsymbol{\rho}^{(2)}$ learnt using the 2 data sets $\mathbf{D}_1$ and $\mathbf{D}_2$ respectively.
Here we keep in mind the possibility that these above factors may not necessarily be mutually independent.

We begin by examining if the assumption of an isotropic phase space that was used in the above Bayesian methodology can result in the recovered inconsistency in the $\boldsymbol{\rho}$ learnt using the 2 data sets. This question is addressed by testing for the support in each of the used data sets, towards the assumption of an isotropic phase space, using a new test of hypothesis that we present here.

As there are two data sets - the relative support in which towards isotropic models are sought - we define two null hypotheses, such that the i-th null, H_i, proposes that the i-th data set $\mathbf{D}_i$ is drawn from an isotropic phase space pdf $\Psi_i(\epsilon)$, $i = 1, 2$. Here $\Psi_i(\epsilon)$ is some real-valued function of the normalised energy $\epsilon \in [0, 1]$, such that $\Psi_i(\epsilon) \geq 0$ if $\epsilon \leq 0$ and $\Psi_i(E) = 0$ otherwise. We refer to a general isotropic phase space pdf as dependent solely on energy ϵ (see Equation 3.8) since we have established in Section 3.1 that isotropic phase space pdfs are functions energy alone. Then under the null that $\mathbf{D}_i$ is sampled from $\Psi_i(\epsilon)$, $\Psi_i(\epsilon) = \mathbf{f}_i$, where $\mathbf{f}_i$ the phase space pdf learnt using data $\mathbf{D}_i$ in the isotropy-assuming Bayesian methodology CHASSIS. Thus,

$$H_i : \mathbf{f}_i = \Psi_i(\epsilon), \quad i = 1, 2, \tag{4.1}$$

for each $i = 1, 2$.

Here we are not making any pre-emptive assumptions about the concurrence of the phase space densities that the two data sets are drawn from. If of course $f_1(\mathbf{x}, \mathbf{v})$ and $f_2(\mathbf{x}, \mathbf{v})$ are isotropic and coincide with each other, then $\Pr(H_1|\mathbf{D}_1)$ and $\Pr(H_2|\mathbf{D}_2)$ will be similar and both $\Pr(H_1|\mathbf{D}_1)$ and $\Pr(H_2|\mathbf{D}_2)$ will be high. If $f_1(\mathbf{x}, \mathbf{v})$ and $f_2(\mathbf{x}, \mathbf{v})$ coincide but neither is an isotropic function of $\mathbf{x}$ and $\mathbf{v}$, $\Pr(H_1|\mathbf{D}_1)$ and $\Pr(H_2|\mathbf{D}_2)$ will be similar but neither $\Pr(H_1|\mathbf{D}_1)$ nor $\Pr(H_2|\mathbf{D}_2)$ will be high. If $f_1(\mathbf{x}, \mathbf{v})$ and $f_2(\mathbf{x}, \mathbf{v})$ do not coincide but are both isotropic, $\Pr(H_1|\mathbf{D}_1)$ and $\Pr(H_2|\mathbf{D}_2)$ will both be high. If $f_1(\mathbf{x}, \mathbf{v})$ and $f_2(\mathbf{x}, \mathbf{v})$ do not coincide and only one is an isotropic function, it is likely that either $\Pr(H_1|\mathbf{D}_1)$ or $\Pr(H_2|\mathbf{D}_2)$ will be high and the other low. Given this framework, we elaborate on the motivation for a new test, developed to work in this context of Bayesian non-parametric learning of the unknown functions, subsequent to an examination of the priors available for the hypotheses as well as for the model parameters.

4.1. *Handling priors.* We first test for H_i as supported by $\mathbf{D}_i$; conventionally, this is done by computing the posterior odds $\Pr(H_i|\mathbf{D}_i)/[1 - \Pr(H_i|\mathbf{D}_i)]$ (Goodman, 1999; Kass and Raftery, 1995). It then appears that we should be comparing the posterior odds of the isotropy-assuming model and the model that does not assume isotropy, once given the data set $\mathbf{D}_1$ and then $\mathbf{D}_2$. However, as mentioned in Section 3.1, in this application the posterior probability of model parameter vectors given either data set, under the anisotropic model is unachievable. Hence we compare $\Pr(H_1|\mathbf{D}_1)/[1 - \Pr(H_1|\mathbf{D}_1)]$ and $\Pr(H_2|\mathbf{D}_2)/[1 - \Pr(H_2|\mathbf{D}_2)]$ instead. This would entail computation of the Bayes factors $\Pr(\mathbf{D}_1|H_1)/[1-\Pr(\mathbf{D}_1|H_1)]$ and $\Pr(\mathbf{D}_2|H_2)/[1-\Pr(\mathbf{D}_2|H_2)]$ and the prior odds $\Pr(H_1)/[1-\Pr(H_1)]$ and $\Pr(H_2)/[1 - \Pr(H_2)]$.

There is no apriori physical reason to constrain the prior $\Pr(H_1)$ as equal to $\Pr(H_2)$. In allowing for such flexibility, we are allowing for the system phase space structure to be characterised by at least two volumes, the state of isotropy in which are not necessarily the same. Since in complex dynamical

systems phase spaces characterised by such insulated pockets are possible, as arising out of non-linear dynamical mechanisms, we do not ignore such a possibility.

Next we ask if there is anything in the astronomical literature to suggest the value of $\Pr(H_i)/[1-\Pr(H_i)]$ $i=1,2$ that is needed in the computation of the Bayes factors. Model-driven numerical simulations of galactic phase spaces have been undertaken (Dekel et al., 2005), though it is precisely because consensus over the status of isotropy in individual real galaxies has not been achieved by such numerical simulations, that the aforementioned mass-anisotropy degeneracy continues to worry the astronomy community. In contrast to such simulations, Côté et al. (2001) discuss isotropy in phase spaces that observed samples of GCs are drawn from, using data from multiple sources that are available for the galactic system in question. However, extrapolating these suggestions to the particular instance of the galaxy of interest in our work (NGC 3379) is risky since the details of the non-linear dynamics responsible for designing the phase space of a galaxy, is particular to the evolution (and structure) of that galaxy. Moreover, the data from the multiple sources that Côté et al. (2001) used is not available for NGC 3379 and the data for this galaxy includes measurements of PNe that live in this galaxy, not just GCs that Côté et al. (2001) considered. Douglas et al. (2007) do indeed discuss the issues regarding isotropy in the phase space that the observed sample of PNe in the galaxy NGC 3379 are drawn from, though they do not provide estimates of the probability of this phase space being isotropic. In other words, the work by Douglas et al. (2007) cannot be employed to constrain $\Pr(H_1)$. The generic simulations of Dekel et al. (2005), will have to be considered to be applicable to the case of this particular galaxy, to suggest $\Pr(H_1)/1-\Pr(H_1)$. However, it is important to note that in their numerical modelling, Dekel et al. (2005) consider the galactic phase space to be a monolithic structure, i.e. they do not consider the possibility that distinct samples of different types of galactic particles could live in insulated volumes of the galactic phase space. Thus, using these simulation-based studies as suggestions for priors do not allow for general models.

The situation here is typical of real-life cases when little information is available towards the prior odds. Thus, we can only suggest $\Pr(H_i)=1-\Pr(H_i)$=0.5, so that the prior odds is unity.

4.2. *Difficulty in computing Bayes factors here.* In the implementation of the posterior odds $\Pr(H_i|\mathbf{D}_i)/[1-\Pr(H_i|\mathbf{D}_i)]$, i=1,2, the computation of the Bayes factors is rendered difficult in light of improper priors on the system parameter vector $(\mathbf{f}^{(i)},\boldsymbol{\rho}^{(i)})^T$. Of course, this is not relevant to applications where priors are not improper. Here, priors $\pi_0(\rho_h^{i)})$ and $\pi_0(\mathbf{f}_j^{(i)})$ are uniform - but over two distinct intervals in values of the state space parameter vector; $j=1,\ldots,N_{eng}$, $h=1,\ldots,N_x$.

Various resolutions of this problem of improper priors has been suggested, including posterior Bayes factors (Aitkin, 1991), intrinsic Bayes factors (Berger and Pericchi, 1996a,b) and fractional Bayes factors (O'Hagan, 1995). In line with the discussion in Section 1, the precise difficulties in translating these suggestion to the application at hand are enumerated below.

(i) The Bayesian learning of $(\mathbf{f}^{(i)},\boldsymbol{\rho}^{(i)})^T$, as delineated above, is possible only if the data distribution is such that there is at least one measured value of v_3 available in each of the N_x bins that the chosen interval of r is binned into, otherwise the method wrongly interpolates linearly between two neighbouring bins belying the non-linearity inherent in the system.

(ii) However, this worry might appear mitigated in light of the suggestion that the training data sets are not required to be real data but can be imaginary (Berger and Pericchi, 2001, 2004; Cano and Salmeron, 2013; Fouskakis, Ntzoufrasy and Draper, 2012; Perez and Berger, 2002). Thus, $\mathbf{D}_i^{(\ell)}$ can be sampled from the posterior predictive distribution $\Pr(\mathbf{D}_i^{(\ell)}|\mathbf{D}_i)$ under the null, i.e.

from $\int \Pr(\mathbf{D}_i^{(\ell)}|\mathbf{f}_0^{(i)},\boldsymbol{\rho}_0^{(i)})\pi(\mathbf{f}_0^{(i)},\boldsymbol{\rho}_0^{(i)}|\mathbf{D}_i)d\boldsymbol{\rho}_0^{(i)}d\mathbf{f}_0^{(i)}$. (Here we adopt the notation that model parameter vectors with subscript "1" refer to those learnt under the more complex model, i.e. the alternative, while those with subscript "0" refer to vectors learnt under the null). The training data $\mathbf{D}_i^{(\ell)}$ thus generated, could then be used to train the prior under the complex model (the model in which the simple model or the null is nested), to be then input as the prior in the computation of the predictive distribution of the i-th data set under the complex model. In other words, we then compute the prior under the complex model as $\Pr(\mathbf{f}_1^{(i)},\boldsymbol{\rho}_1^{(i)}) = \int \pi(\mathbf{f}_1^{(i)},\boldsymbol{\rho}_1^{(i)}|\mathbf{D}_i^{(\ell)})d\mathbf{D}_i^{(\ell)}$ and subsequently input this to compute the predictive distribution under the complex model, $m_1(\mathbf{D}_i^{(\ell)}) = \int \Pr(\mathbf{D}_i^{(\ell)}|\mathbf{f}_1^{(i)},\boldsymbol{\rho}_1^{(i)})\Pr(\mathbf{f}_1^{(i)},\boldsymbol{\rho}_1^{(i)})d\mathbf{f}_1^{(i)}d\boldsymbol{\rho}_1^{(i)}$, which would be invoked in the Bayes factor. However, in our work, it is not possible to compute the posterior probability of the model parameter vectors under the complex model that does not assume isotropy but incorporates a general anisotropic phase space pdf within its definition. In other words, the posterior probability of $\mathbf{f}_1^{(i)}$ and $\boldsymbol{\rho}_1^{(i)}$ though perceived, is not computable under a general anisotropic phase space pdf (see Section 3.1).

(iii) This shortcoming notwithstanding, imaginary training data can in principle be implemented to give the posterior odds of the null to its complement, given a data set. However, the computational intricacy involved in averaging over all possible imaginary samples is formidable - such averaging is required to compute the "expected-posterior prior" of the model parameter vectors under the null, i.e. $\int \pi(\mathbf{f}_0^{(i)},\boldsymbol{\rho}_0^{(i)}|\mathbf{D}_i^{(\ell)})m_0(\mathbf{D}_i^{(\ell)})d\mathbf{D}_i^{(\ell)}$ (Fouskakis, Ntzoufrasy and Draper, 2012). We would then need to generate a large sample of training data sets, and for each these training data sets, we would need to learn the unknown model parameter vector under the assumption of isotropy. This suggests running as many long MCMC chains to convergence, as there are training data sets that the posterior of the model parameter vectors for the null, is averaged over, in the numerical computation of the expected-posterior prior under the null. This is required to be a large number, if the expected non-linearity in the joint posterior probability of all the unknown parameters is to be adequately explored. Given such a computationally intensive method, we seek an new method that is numerically less cost intensive.

In view of this, we present a new distribution-free test of hypothesis that works by taking the full posterior structure of the model parameters into account. This test is inspired by the Fully Bayesian Significance Test (FBST) advanced by de B. Pereira and Stern (1999); de B. Pereira, Stern and Wechsler (2008).

5. Method. In our test, a measure of the support in the data $\mathbf{D}_i$ for the null H_i, i.e. $\Pr(H_i|\mathbf{D}_i)$ is given by the complement of the integral of the posterior probability density of the model parameter vector $(\mathbf{f}^{(i)},\boldsymbol{\rho}^{(i)})^T$ given data $\mathbf{D}_i$, over all those parameters, the posterior of which exceeds the maximal posterior under the null. In fact, it is a rephrasing of this integral that allows us to compute $\Pr(H_i|\mathbf{D}_i)$ in this context of non-parametric multimodal inference in high-dimensions.

In this section we first describe FBST, then introduce this new test and then proceed to describe its adaptation for our application.

5.1. *FBST.* The Fully Bayesian Significance Test (FBST) presented by de B. Pereira, Stern and Wechsler (2008), tests the sharp null hypothesis that the relevant parameter, θ, has a value θ_0, i.e. $H_0 : \theta = \theta_0$.

Here θ is assumed to be distributed continuously in the parameter space Θ. In FBST, the evidence in favour of H_0 is measured by first identifying the parameter $\theta^\star$ for which posterior probability under the null, given data $\mathbf{D}$, is a maximum, i.e. $\theta^\star = \arg\max_{H_0} \Pr(\theta|\mathbf{D})$. de B. Pereira, Stern and Wechsler (2008) suggest the evidence in favour of H_0 to be:

$$ev = 1 - \Pr(\theta \in T|\mathbf{D}), \quad \text{where} \tag{5.1}$$

$$T = \{\theta : \Pr(\theta|\mathbf{D}) > \Pr(\theta^\star|\mathbf{D})\}, \tag{5.2}$$

The intuitive justification of the above is the following. Higher is the probability of achieving θ that is more consistent with the observed data than is maximally achieved under the null, less is the support in this data for the null; such θ live in the set T by definition (Equation 5.2). Thus, higher is the probability of identifying $\theta \in T$, the less is the support in $\mathbf{D}$ for H_0. Then FBST quantifies support in the data against the null hypothesis as the integral of the posterior probabilities of θ, for $\theta \in T$. Thus, FBST involves identification of $\theta^\star$ via optimisation, followed by integration over T. One aspect of this test is that the measure of evidence in favour of the null obeys the Likelihood Principle (Basu, 1975; Berger and Wolpert, 208; Birnbaum, 1962).

From this discussion, it may appear that the way the evidence value ev is defined above, it is not invariant to re-parametrisation of the parameter space. This problem has been addressed by Madruga, Pereira and Stern (2003) and de B. Pereira, Stern and Wechsler (2008) via the suggestion that the formulation of the ev be in terms of the surprise function $s(\cdot)$ relative to a reference density $r(\cdot)$, where $s(\theta) := \dfrac{\Pr(\theta|\mathbf{D})}{r(\theta)}$, where the reference density on θ is $r : \Theta \longrightarrow \mathbb{R}$. The purpose of the normalisation of the posterior of θ given the data, by the reference density function $r(\theta)$ is to ensure that for any transformation of θ, such as $\omega = \Xi(\theta)$, the supremum of the transformed surprise function remains invariant, i.e. the measure of evidence in favour of the null, $\bar{ev}$, is rendered invariant to re-parametrisation (shown by Madruga, Pereira and Stern, 2003, for a bijective and continuously differentiable $\Xi(\cdot)$). In this revised interpretation, the FBST procedure involves identifying $\theta^\star = \arg\max_{H_0} s(\theta)$ and the evidence in favour of the null is

$$\bar{ev} = 1 - \Pr(\theta \in \bar{T}|\mathbf{D}), \quad \text{where}$$

$$\bar{T} = \{\theta : s(\theta) > s(\theta^\star)\}, \tag{5.3}$$

de B. Pereira, Stern and Wechsler (2008); Madruga, Pereira and Stern (2003) suggest possible interpretations of $r(\theta)$. If $r(\theta)$ is treated as the uniform density $U(\theta)$ in FBST, this effectively equates $\bar{ev}$ to ev (Equation 5.2).

5.2. *The new test.* In the non-parametric test of hypothesis that is presented here, the support in the data $\mathbf{D}_i$ for the null hypothesis H_i is given as

$$\Pr(H_i|\mathbf{D}_i) = 1 - \int_{T_i} \pi(\boldsymbol{\theta}|\mathbf{D}_i) d\boldsymbol{\theta}, \quad \text{where} \tag{5.4}$$

$$T_i = \left\{ \boldsymbol{\theta} : \frac{\pi(\boldsymbol{\theta}|\mathbf{D}_i)}{r(\boldsymbol{\theta})} > \frac{\pi_{H_i}(\boldsymbol{\theta}^\star|\mathbf{D}_i)}{r(\boldsymbol{\theta})} \right\}, \tag{5.5}$$

$i = 1, 2$, and $\pi(\boldsymbol{\theta}|\mathbf{D}_i)$ is the posterior probability density of the model parameter vector $\boldsymbol{\theta} = (f_1, f_2, \ldots, f_{N_{eng}}, \rho_1, \rho_2, \ldots, \rho_{N_x})^T$ given data $\mathbf{D}_i$. $\pi_{H_i}(\boldsymbol{\theta}|\mathbf{D}_i)$ is the posterior probability density of

$\boldsymbol{\theta}$ under the null H_i. To ensure invariance to reparametrisation of the parameter space, we define the set T_i in terms of the surprise function $s(\boldsymbol{\theta}) := \dfrac{\pi(\boldsymbol{\theta}|\mathbf{D}_i)}{r(\boldsymbol{\theta})}$, where we choose to work with a reference density $r(\boldsymbol{\theta})$ that is uniform in $\boldsymbol{\theta}$. Here T_i is defined in Equation 5.5 as the set of those values of the parameter $\boldsymbol{\theta}$, the posteriors of which exceed the maximal posterior achieved under the null H_i. Thus, T_i consists of those parameter values that are even more consistent with data $\mathbf{D}_i$ than is $\boldsymbol{\theta}^\star$, which is the most consistent value under the null. In other words, the value of the model parameter that maximises the posterior under the null, $\pi_{H_i}(\boldsymbol{\theta}|\mathbf{D}_i)$, is identified as $\boldsymbol{\theta}^\star$, i.e. $\boldsymbol{\theta}^\star = \arg\left[\max_{\Theta_i} \pi_{H_i}(\boldsymbol{\theta}|\mathbf{D}_i)\right]$. Having identified $\boldsymbol{\theta}^\star$, we partition the parameter space Θ_i into the subspaces T_i and $\bar{T}_i$, where T_i is defined as in Equation 5.5. Thus,

$$\begin{array}{rcl} \Theta_i & = & T_i \cup \bar{T}_i \\ T_i = \{\boldsymbol{\theta} : \pi(\boldsymbol{\theta}|\mathbf{D}_i) > \pi_{H_i}(\boldsymbol{\theta}^\star|\mathbf{D}_i)\} & \Longleftrightarrow & \bar{T}_i = \{\boldsymbol{\theta} : \pi(\boldsymbol{\theta}|\mathbf{D}_i) \leq \pi_{H_i}(\boldsymbol{\theta}^\star|\mathbf{D}_i)\} \end{array}$$

We realise that larger is the proportion of parameter values that are more consistent with the data than what is maximally achievable under the null, the less is the support in the data for the null. Here, "consistency with the data" is measured by the posterior probability density of parameter vector $\boldsymbol{\theta}$ given data $\mathbf{D}_i$. Then posterior probability density of all $\boldsymbol{\theta} \in T_i$ given $\mathbf{D}_i$ contribute to support in $\mathbf{D}_i$ against the null. In other words, $\pi(\boldsymbol{\theta}|\mathbf{D}_i)$ integrated over T_i defines total support against the null as $1 - \Pr(H_i|\mathbf{D}_i)$. This gives Equation 5.5. So,

$$\begin{array}{rcl} 1 - \Pr(H_i|\mathbf{D}_i) & = & \displaystyle\int_{T_i} \pi(\boldsymbol{\theta}|\mathbf{D}_i) d\boldsymbol{\theta} \quad \text{or} \\ \Pr(H_i|\mathbf{D}_i) & = & \displaystyle 1 - \int_{T_i} \mathcal{L}(\boldsymbol{\theta}|\mathbf{D}_i)\pi_0(\boldsymbol{\theta}) d\boldsymbol{\theta} \end{array} \tag{5.6}$$

Thus, unlike Bayes factors - the computation of which involves integrating over the parameter space - this test involves integrating over the subspace T_i.

Now, the posterior probability density of $\boldsymbol{\theta}$ given $\mathbf{D}_i$, integrated over the subspace T_i, is the probability of $\boldsymbol{\theta}$ to live in this subspace, given the data, i.e.

$$\Pr(\boldsymbol{\theta} \in T_i|\mathbf{D}_i) = \int_{T_i} \pi(\boldsymbol{\theta}|\mathbf{D}_i) d\boldsymbol{\theta} = 1 - \Pr(H_i|\mathbf{D}_i). \tag{5.7}$$

Then the integral of $\pi(\boldsymbol{\theta}|\mathbf{D}_i)$ over subspace T_i is approximated by the proportion of $\boldsymbol{\theta} \in T_i$; this proportion is determined empirically. This crucially eases the computational burden in the context of this non-parametric inference in high-dimensions.

5.3. *Motivation for a Bayesian test.* Could a frequentist test have worked for our purpose? That would require us to define a test statistic S and to aim at finding the probability that S is more or equally unlikely than the value of this statistic, given the data. A possible candidate for S could be $S := 1/\mathcal{L}(\boldsymbol{\beta}|\mathbf{D}_i)$ where $\mathcal{L}(\boldsymbol{\beta}|\mathbf{D}_i)$ is the likelihood of the "relevant parameters" $\boldsymbol{\beta}$ given data $\mathbf{D}_i$. Such a choice of S is motivated to ultimately allow us to reject or not reject (rather than accept or not accept) the null within the paradigm of a non-parametric frequentist test. Now $\boldsymbol{\beta}$, the "parameter of relevance" to this null must measure how isotropic the phase space *pdf* is, given that the null states: $\mathbf{f}_i$ learnt using $\mathbf{D}_i$ is isotropic (Equation 4.1). The concept of a parametrising how isotropic (or anisotropic) the

phase space pdf is, exists in the astronomy literature in the form of the anisotropy parameter (discussed above in Section 3.1). However, in general there is no measurement available that allows for the quantification of this anisotropy parameter (Binney and Tremaine, 1987), leading to the worry referred to as the mass-anisotropy degeneracy (Côté et al., 2001, 2003; Koopmans, 2006; Łokas and Mamon, 2003). In fact, in lieu of such information, in the community, the anisotropy parameter is arbitrarily assigned values, and for each such assignment, a new phase space pdf (and $\rho(r)$) is obtained from the data (see Binney and Tremaine, 1987; Douglas et al., 2007, for examples of this). In other words, $\mathcal{L}(\boldsymbol{\beta}|\mathbf{D}_i)$ cannot be achieved. Even for other prescriptions of the test statistic S, it is bound to bear dependence on how likely the parameters of relevance are, given the data. However, the computation of this likelihood is impossible given the nature of the available information, i.e. S is fundamentally not achievable in this application. Above, $i = 1, 2$.

In lieu of a way to compute the likelihood of $\boldsymbol{\beta}$, we compute the probability of $\pi(\boldsymbol{\theta}|\mathbf{D}_i) \leq \pi_{H_i}(\boldsymbol{\theta}^*|\mathbf{D}_i)$, (Equation 5.6) to subsequently compute $\Pr(H_i|\mathbf{D}_i)$ (Equation 5.7). Thus, in this framework, we can work in terms of the model parameter vector $\boldsymbol{\theta}$ rather than the parameter $\boldsymbol{\beta}$ that measures anisotropy of the phase space. Attempting to compare the likelihood of $\boldsymbol{\theta}$ given $\mathbf{D}_i$, with the maximum value of this likelihood under the null, is synonymous to comparing $\Pr(\mathbf{D}_i|\boldsymbol{\theta})$ with the maximum value of this probability under the null. However, by definition, the likelihood considers variation in data given a $\boldsymbol{\theta}$ while we need to find the probability of the null given data set, i.e. such a comparison cannot lead to $\Pr(H_0|\mathbf{D}_i)$. Thus, one needs to work in the Bayesian framework.

Thus, we prefer to work in parameter space and at the same time, given the radically different sample sizes of the GC and PNe data, it is imperative that the test be independent of sample-size effects. We recall that the sample size of the PNe data is about 5.6 times the size of the GC data.

For the above reasons, a (non-parametric) frequentist test cannot be useful in this application.

5.4. *Implementation of the new test.* In contrast to the parametric FBST discussed by de B. Pereira, Stern and Wec (2008), in our work, the the posterior of the system variable, given the data, does not conform to any parametric form. Also, to render implementation possible in this context, we avoid a direct computation of the integral of the posterior probability density of the model parameter vector given the data at hand. Given this, we present our test below.

The null hypothesis H_i that we aim to test for, is given above in Equation 4.1, $i = 1, 2$. Both H_1 and H_2 represent sharp hypotheses. In our work, the model parameter vector variable is $\boldsymbol{\theta} = (f_1, f_2, \ldots, f_{N_{eng}}, \rho_1, \rho_2, \ldots, \rho_{N_x})^T \in \mathcal{F} \times \mathcal{R}$. For $\mathbf{f}^{(i)}$ and $\boldsymbol{\rho}^{(i)}$ learnt using the i-th data set $\mathbf{D}_i$, we define

$$\boldsymbol{\theta}^{(i)} := ((\mathbf{f}^{(i)})^T, (\boldsymbol{\rho}^{(i)})^T)^T \quad i = 1, 2.$$

5.4.1. *Identification of posterior-maximising model parameter vector, under the null.* With the aim of identifying the subspace T_i, ($i = 1, 2$), we identify the $\boldsymbol{\theta}^{(i)}$ vector that maximises the posterior $[\mathbf{f}^{(i)}, \boldsymbol{\rho}^{(i)}|\mathbf{D}_i]$ under the null, given the data. This posterior maximising, null abiding vector is referred to as $\boldsymbol{\theta}^{(i\star)}$. In order to identify this vector, the following construct is used.

- During the posterior inference on $\boldsymbol{\theta}^{(i)}$, performed with adaptive Metropolis-Hastings, let the current state vector be $\boldsymbol{\theta}_j^{(i)}$, in the j-th iteration, $j = 1, \ldots, N_0$, where the chain is N_0 steps long. Upon convergence, the unknown $\mathbf{f}^{(i)}$ and $\boldsymbol{\rho}^{(i)}$ are learnt within 95% HPD credible regions. From a given chain, the model parameter vector $\boldsymbol{\theta}^{(Mi)} := ((\mathbf{f}^{(Mi)})^T, (\boldsymbol{\rho}^{(Mi)})^T)^T$ corresponding to the mode of the posterior $[\mathbf{f}^{(i)}, \boldsymbol{\rho}^{(i)}|\mathbf{D}_i]$ is identified.

- Then from the identified $\mathbf{f}^{(Mi)}$, at the identified $\boldsymbol{\rho}^{(Mi)}$, we simulate an N_i-sized data set of the observable variables, i.e. X_1, X_2 and V_3. Let this generated data set be $\mathbf{D}_{i,\,gen} := \{(X^{(k)}_{1,\,gen}, X^{(k)}_{2,\,gen}, V^{(k)}_{3,\,gen})\}_{k=1}^{N_i}$.
- Importantly, we understand that by construction, this data set is sampled from an isotropic phase space density since $\mathbf{f}^{(Mi)}$ was learnt under the model assumption of isotropy. This data set is simulated from $\mathbf{f}^{(Mi)}$ at $\boldsymbol{\rho}^{(Mi)}$ using rejection sampling, according to the following algorithm.

1. We solve Poisson equation (Equation 3.3) in a spherical geometry to get, $\Phi(r) = \dfrac{-GM(r)}{r}$ where $M(r) = \displaystyle\int_{s=0}^{r} 4\pi\rho(s)s^2 ds$ and G is a known constant. Then discretising this integral, we define

$$\begin{aligned} M(r) &= \sum_{q=1}^{p} \frac{4\pi}{3}[q^3\delta^3 - (q-1)^3\delta^3]\rho_q + \frac{4\pi}{3}[r^3 - p^3\delta^3]\rho_{p+1} \quad \text{for} \quad r \in [p\delta, (p+1)\delta), \\ M(r) &= \sum_{q=1}^{N_x} \frac{4\pi}{3}[q^3\delta^3 - (q-1)^3\delta^3]\rho_q \quad \text{for} \quad r \geq N_x\delta, \\ M(r) &= \frac{4\pi}{3}[r^3\delta^3]\rho_1 \quad \text{for} \quad 0 \leq r \leq \delta, \end{aligned} \tag{5.8}$$

   where $1 \leq p \leq N_x - 1$. Here $N_x\delta$ is the maximum radius to which data are available and ρ_q is the gravitational mass density in the q-th radial bin, i.e. $\rho(r) = \rho_q$ if $r \in [(q-1)\delta, q\delta)$, $q = 1, \ldots, N_x$ (introduced in Section 3.2). This defines $\Phi(r)$ for any $r \geq 0$, given the identified $\boldsymbol{\rho}^{(Mi)}$. For the $\boldsymbol{\rho}^{(Mi)}$ learnt using $\mathbf{D}_i^{(gen)}$, we compute $\Phi(r)$ as above.
2. At this step, we sample the normalised value of the energy function, ϵ defined in Section 3.2. As this normalised value $\in [0, 1]$, we choose ϵ randomly from $\mathcal{U}[0, 1]$, where Here $\mathcal{U}[a, b]$ is the uniform distribution over the range $[a, b]$, for any $a, b \in \mathbb{R}$. Let the chosen ϵ be such that it lies in the t-th energy bin, i.e. $\epsilon \in [(t-1)\delta_E, t\delta_E]$, $t = 1, 2, \ldots, N_{eng}$. Over this t-th energy bin, let the component of $\mathbf{f}^{(Mi)}$ be $f_t^{(Mi)}$.
3. The 3 components of the velocity vector, V_1, V_2, V_3 are continuous in $[-\sqrt{-2\Phi(0)}, \sqrt{-2\Phi(0)}]$, while the components of the spatial vector are continuous in $[-N_x\delta, N_x\delta]$. We set, $X_1, X_2, X_3 \sim \mathcal{U}[-N_x\delta, N_x\delta]$ and $V_1, V_2, V_3 \sim \mathcal{U}[-\sqrt{-2\Phi(0)}, \sqrt{-2\Phi(0)}]$.
4. Draw values of X_1, X_2, X_3 from $\mathcal{U}[-N_x\delta, N_x\delta]$. Using these chosen values x_1, x_2, x_3, we obtain the value of the spherical radius $r = \sqrt{x_1^2 + x_2^2 + x_3^2}$. Let r be such that it lies in the q-th radial bin, i.e. $r \in [(q-1)\delta, q\delta]$, $q = 1, 2, \ldots, N_x$. For this chosen r, we then compute $\Phi(r)$ using $M(r)$ from Equation 5.8 and the definition $\Phi(r) = \dfrac{-GM(r)}{r}$. We normalise $\Phi(r)$ by $\Phi(0)$, so that $\Phi(r)$ now lives in the range $[0, 1]$.
5. Check if the chosen $\epsilon > \Phi(r)$. If not, go back to step number 2. If yes, then define the magnitude of the velocity vector $\boldsymbol{V}$ as $\sqrt{[2(\epsilon - \Phi(r))]}$, where $\boldsymbol{V} = (V_1, V_2, V_3)^T$. Now, $V_1, V_2, V_3 \sim \mathcal{U}[-\sqrt{-2\Phi(0)}, \sqrt{-2\Phi(0)}]$. So we draw v_1, v_2, v_3 individually from this uniform distribution.
6. In this step, we sample from $f_t^{(Mi)}$ using rejection sampling. Here the chosen ϵ is in the t-th energy-bin so that $f_t^{(Mi)}$ is the value of $f(\epsilon)$ in our discretised model for the chosen

ϵ. The rejection sampling is done by checking if $\frac{f_t^{(Mi)}}{g(\epsilon)} > u$ or not, where u is a random number in $[0,1]$, $u \sim \mathcal{U}[0,1]$. Here $g(\epsilon)$ is the proposal density function that is chosen to envelope over $f(\epsilon)$, $\forall\epsilon$, and is defined as $g(\epsilon) = 1.05\forall\epsilon$. This is an adequate choice because the phase space *pdf* $f(\epsilon)$ is normalised to be in $[0,1]$. If the above inequality holds, we allow an integer-valued flag, γ, an increment of 1 and accept the values x_1, x_2 and v_3 as chosen in steps (4) and (5) respectively, as the γ-th line in $\mathbf{D}_i^{(gen)}$. We iterate over points 2 to 6, until γ equals N_i.

- We input $\mathbf{D}_i^{(gen)}$ to CHASSIS, to start a new chain. Post burn-in, samples of $\boldsymbol{\theta}^{(i)}$ vectors generated in each iterative step are recorded. It is to be noted that any such recorded $\boldsymbol{\theta}^{(i)}$ abides by the null hypothesis H_i since it is recovered using the input data set $\mathbf{D}_i^{(gen)}$ that is drawn from an isotropic phase space density function - one that is learnt from the observed data set $\mathbf{D}_i$ under the assumption that the phase space is isotropic. In this recorded sample of vales of $\boldsymbol{\theta}^{(i)}$, that which maximises the posterior $[\mathbf{f}^{(i)}, \boldsymbol{\rho}^{(i)}|\mathbf{D}_i^{(gen)}]$, is the null-abiding, posterior-maximising parameter

$$\boldsymbol{\theta}^{(i\star)} := ((\mathbf{f}^{(i\star)})^T, (\boldsymbol{\rho}^{(i\star)})^T))^T.$$

5.4.2. *Probability of membership in subspace T_i.* We need to identify the sub-space T_i in which live model parameter vectors, the posterior of which exceeds the posterior probability density $\pi(\boldsymbol{\theta}^{(i\star)}|\mathbf{D}_i^{(gen)})$. We are required to integrate the posterior probability density of such model parameter vectors, over all parameters that live in the subspace T_i. Thus, this integral is equal to $\Pr(\boldsymbol{\theta} \in T_i|\mathbf{D}_i)$. Thus, implementation is possible in this instance of non-parametric, multimodal inference in a high-dimensional state space, by approximating this probability of membership in T_i with a case counting scheme in which, probability of the membership of $\boldsymbol{\theta}$ in T_i computed, conditioned on the observed data $\mathbf{D}_i$. This is the proportion of the model parameter vectors for which $\pi(\boldsymbol{\theta}|\mathbf{D}_i) > \pi(\boldsymbol{\theta}^{(i\star)}|\mathbf{D}_i^{(gen)})$, as recovered in the post-burnin stage of chains run with real data $\mathbf{D}_i$.

Thus, let there be a total of $A_i \in \mathbb{Z}_+$ number of samples of $\boldsymbol{\theta}$ vectors recovered in the post-burnin stage in these chains run with data $\mathbf{D}_i$. Out of these, let B_i number of $\boldsymbol{\theta}$ vectors be such that $\pi(\boldsymbol{\theta}^{(i)}|\mathbf{D}_i) > \pi(\boldsymbol{\theta}^{(i\star)}|\mathbf{D}_i^{(gen)})$. Here, $B \in \mathbb{Z}_+$, $B_i \leq A_i$. Then the fraction B_i/A_i is an approximation to the probability that $\boldsymbol{\theta} \in T_i$, given $\mathbf{D}_i$. Then recalling Equation 5.7, we state

$$\Pr(H_i|\mathbf{D}_i) = 1 - \frac{B_i}{A_i}, \tag{5.9}$$

i=1,2. We compute $\Pr(H_1|\mathbf{D}_1)$ and $\Pr(H_2|\mathbf{D}_2)$, and compare the two.

6. Support for the assumption of isotropy of synthetic phase spaces that data sets are simulated from. In this section, we implement the new test to quantify support for the assumption of isotropy, in data that have been sampled from synthetic phase spaces. We choose an isotropic phase space density function $f_{iso}(\mathbf{x}, \mathbf{v})$ and an anisotropic phase space *pdf* $f_{aniso}(\mathbf{x}, \mathbf{v})$, and sample 2 data sets $\mathbf{D}_{iso}$ and $\mathbf{D}_{aniso}$ from these phase space density functions, respectively. We choose the sample size of $\mathbf{D}_{aniso}$ to be 54 while the sample size of $\mathbf{D}_{iso}$ is chosen to be 270, i.e. 5 times that of $\mathbf{D}_{aniso}$. These numbers mimic the ratio of sample sizes of the real data - ratio of size of $\mathbf{D}_1$ to that of $\mathbf{D}_2$ is about 5.6. The

chosen parametric forms of the phase space density functions are

$$
\begin{aligned}
f_{iso}(\mathbf{x},\mathbf{v}) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{\epsilon}{2\sigma^2}\right), \\
f_{aniso}(\mathbf{x},\mathbf{v}) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{\epsilon}{2\sigma^2}\right)\exp\left(-\frac{|\mathbf{L}|^2}{r_a^2\sigma^2}\right) \\
\text{where}\quad \epsilon = \frac{\mathbf{v}^2/2+\Phi(r)}{\Phi_0}, &\qquad \mathbf{L}=\mathbf{x}\wedge\mathbf{v}.
\end{aligned} \tag{6.1}
$$

Here, we choose the anisotropy scale $r_a = 4$ in units of kpc (the astronomical unit of length at galactic scales) and $\sigma = 220$ in units of km s^{-1}. The motivation for these numerical values is not strong; familiarity with length and velocity scales similar to those in the Milky Way is the only guiding factor. The chosen form of $\Phi(r)$ is maintained as same in both definitions of the synthetic phase space pdfs, as that of the Plummer model (Binney and Tremaine, 1987),

$$
\Phi(r) = -\frac{GM_0}{\sqrt{r_c^2+r^2}}, \tag{6.2}
$$

where we chose the parameters to be $M_0 = 4\times 10^{11}$ times the mass of the Sun and r_c=8 kpc. It may be noted that M_0 and r_c represent the total gravitational mass of the synthetic system and the scale length of the Plummer model. Again, choices for the numerical values of M_0 and r_c are motivated by nothing more than familiarity with such values that describe our galaxy. G is a known physical constant, the universal gravitational constant.

Having defined the true model, we then simulated $\mathbf{D}_{iso}$ and $\mathbf{D}_{aniso}$ from respective phase space densities, where each data set contained information on X_1, X_2 and V_3. The sampled V_3 data was chosen to be characterised by Gaussian noise $\sim\mathcal{N}(0,20^2)$ which is typical of real-life galaxies that are nearby (Douglas et al., 2007). The model parameter vectors $\boldsymbol{\rho}$ and $\mathbf{f}$ learnt by porting $\mathbf{D}_{iso}$ and $\mathbf{D}_{aniso}$ to the algorithm CHASSIS is shown in black in the right and left panels respectively, of Figure 4. The modal values of the learnt components of these model parameter vectors are displayed with open circles while the error-bars in the plots depict the 95% HPDs. Given the lack of information that typifies inference in real-life galaxies, here we used uniform priors $\pi_0(\rho_h) = \dfrac{1}{\rho_h^{(0)}[10^2-10^{-5}]}$ and $\pi_0(f_j)=1$, where $\rho_h^{(0)}$ is the h-th component of the seed value of the gravitational mass density vector. For the implementation of $\mathbf{D}_{iso}$, we used $h=1,\ldots,19$, $j=1,\ldots,9$ and for the implementation of $\mathbf{D}_{aniso}$, we used $h=1,\ldots,14$, $j=1,\ldots,9$.

Thus, the unknowns are learnt under the assumption of phase space isotropy, using data - one of which is, and one of which is not sampled from an isotropic phase space density. We refer to the vectors learnt using $\mathbf{D}_{iso}$ as $\boldsymbol{\rho}_{iso}$ and $\mathbf{f}_{iso}$ and those learnt using $\mathbf{D}_{aniso}$ as $\boldsymbol{\rho}_{aniso}$ and $\mathbf{f}_{aniso}$. In order to check for the support in these 2 data sets for the assumption of phase space isotropy, we first seek the model parameter vector pair $\mathbf{f}^{(i\star)}, \boldsymbol{\rho}^{(i\star)}$ that maximises the posterior under the null, given the generated data set $\mathbf{D}_i^{(gen)}$. Here $\mathbf{D}_i^{(gen)}$ is generated by sampling from the modal parameter vectors $\mathbf{f}^{(Mi)}$ at $\boldsymbol{\rho}^{(Mi)}$, where i stands for "iso" or "$aniso$" (see Section 5.4 for details of this sampling). Here the null $H_i : \mathbf{f}^{(i)} = \Psi_i(\epsilon)$.

In Figure 5, the logarithm of the posterior probability density $\pi(\mathbf{f}^{(iso)}, \boldsymbol{\rho}^{(iso)}|\mathbf{D}_{iso})$ and $\pi(\mathbf{f}^{(aniso)}, \boldsymbol{\rho}^{(aniso)}|\mathbf{D}_{aniso})$ are shown in black in the right and left panels respectively. The maximal posterior probability density under the null in the two cases are $\pi(\mathbf{f}^{(iso)}, \boldsymbol{\rho}^{iso)}|\mathbf{D}_{iso}^{(gen)})$ and

TABLE 2

Table displaying support in synthetic data $\mathbf{D}_{iso}$ and $\mathbf{D}_{aniso}$ (simulated respectively from known isotropic and anisotropic parametric phase space density functions), for the null that the data is sampled from an isotropic phase space density. Column 2 shows the ratio B_i/A_i, i.e. the fractional number of iterations in which posterior probability density of $\mathbf{f}^{(i)}$ and $\boldsymbol{\rho}^{(i)}$, given $\mathbf{D}_i$ exceeds the maximal posterior probability achieved under the null, given generated data $\mathbf{D}_i^{(gen)}$ that has been sampled from the modal $\mathbf{f}^{(Mi)}$ learnt using $\mathbf{D}_i$. Here i stands for "iso" or "$aniso$". Here, the fraction is taken out of the total number of iterations in the chain, post-burnin. Column 3 delineates the probability of the i-th null conditional on the i-th synthetic data.

Simulated data $\mathbf{D}_i^{(\star)}$	A_i/B_i	$\Pr(H_i\|\mathbf{D}_i)$
$\mathbf{D}_{iso}$	0	1
$\mathbf{D}_{aniso}$	1	0

$\pi(\mathbf{f}^{(aniso)}, \boldsymbol{\rho}^{(aniso)}|\mathbf{D}_{aniso}^{(gen)})$, are shown in red, in the right and left panels respectively. Here i stands for "iso" or "$aniso$". The fractional number of MCMC samples $((\mathbf{f}^{(i)})^T, (\boldsymbol{\rho}^{(i)})^T)^T$ in the chain run with $\mathbf{D}_i$, with posterior probability density values that exceed $\pi(\mathbf{f}^{(iso)}, \boldsymbol{\rho}^{iso)}|\mathbf{D}_i^{(gen)})$, is an empirical approximation of the probability for $((\mathbf{f}^{(i)})^T, (\boldsymbol{\rho}^{(i)})^T)^T$ to live in subspace T_i. The complement of this probability gives $\Pr(H_i|\mathbf{D}_i)$ (see Equation 5.7).

The computed support in $\mathbf{D}_i$ for the null H_i is tabulated in Table 2. As is expected, for the data set $\mathbf{D}_{aniso}$ that is simulated from an isotropy-defying phase space density, the support in the data for the null that the data is sampled from an isotropy-assuming phase space pdf, is 0. On the other hand, when the implemented data $\mathbf{D}_{iso}$ is truly sampled from an isotropic distribution, the method recovers a support in the data for the null that is 1. That the inference in being performed in a multimodal setting is brought out in Figure 6 that depicts the multimodal nature of the marginal posterior of the example parameter ρ_6, learnt using the data set $\mathbf{D}_{aniso}$.

7. Support for the assumption of isotropic phase space of a real galaxy. To present the quantification of the support in the real, observed data sets $\mathbf{D}_1$ and $\mathbf{D}_2$ each, towards the assumption of an isotropic phase space pdf, we recall the results of the Bayesian learning of the model parameters using these data sets. These results were presented in Section 3.3.

Our new test, as described above, is invoked to estimate if the assumption of isotropy is supported, in the two different particle samples that we deal with.

The computed support in the two data sets $\mathbf{D}_1$ and $\mathbf{D}_2$ to have been sampled from isotropic phase space pdfs are $\Pr(H_1|\mathbf{D}_1)$ and $\Pr(H_2|\mathbf{D}_2)$ respectively, are represented in Table 3.

The details of the computational and inferential procedure are as follows. The chains are at least 400,000 iterations long, when data $\mathbf{D}_i$ is used to learn $\boldsymbol{\rho}^{(i)}$ and $\mathbf{f}^{(i)}$, $i = 1, 2$, under the assumption of phase space isotropy, in the Bayesian non-parametric method CHASSIS. Uniform priors on $\mathbf{f}$ and $\boldsymbol{\rho}$ are used. A sample $\mathbf{D}_i^{(gen)}$, was drawn from the learnt $\mathbf{f}^{(Mi)}$ at the learnt $\boldsymbol{\rho}^{(Mi)}$, with i=1,2. The chain runs with these generated data $\mathbf{D}_i^1$, is the same length as the original chains.

Comparing the computed $\Pr(H_1|\mathbf{D}_1)$ and $\Pr(H_2|\mathbf{D}_2)$ across the chains implies that the assumption of isotropy is more likely to be invalid for the phase space from which the PNe data $\mathbf{D}_1$ are drawn than from which the GC data $\mathbf{D}_2$ are drawn. Basically, support in real data $\mathbf{D}_1$ for the assumption of isotropy is distinct from that in $\mathbf{D}_2$. This implies that the $f_1(\boldsymbol{x}, \boldsymbol{v}) \neq f_2(\boldsymbol{x}, \boldsymbol{v})$, where phase space pdf that $\mathbf{D}_1$ is sampled from is $f_1(\boldsymbol{x}, \boldsymbol{v})$ and $\mathbf{D}_2 \sim f_2(\boldsymbol{x}, \boldsymbol{v})$. However, both data sets carry information on the phase space coordinates $(\boldsymbol{X}^T, \boldsymbol{V}^T)^T$ of the same galaxy, i.e. both data sets are sampled from pdfs that describe the phase space structure of all or some volume inside the same galactic phase space $\mathcal{W}$.

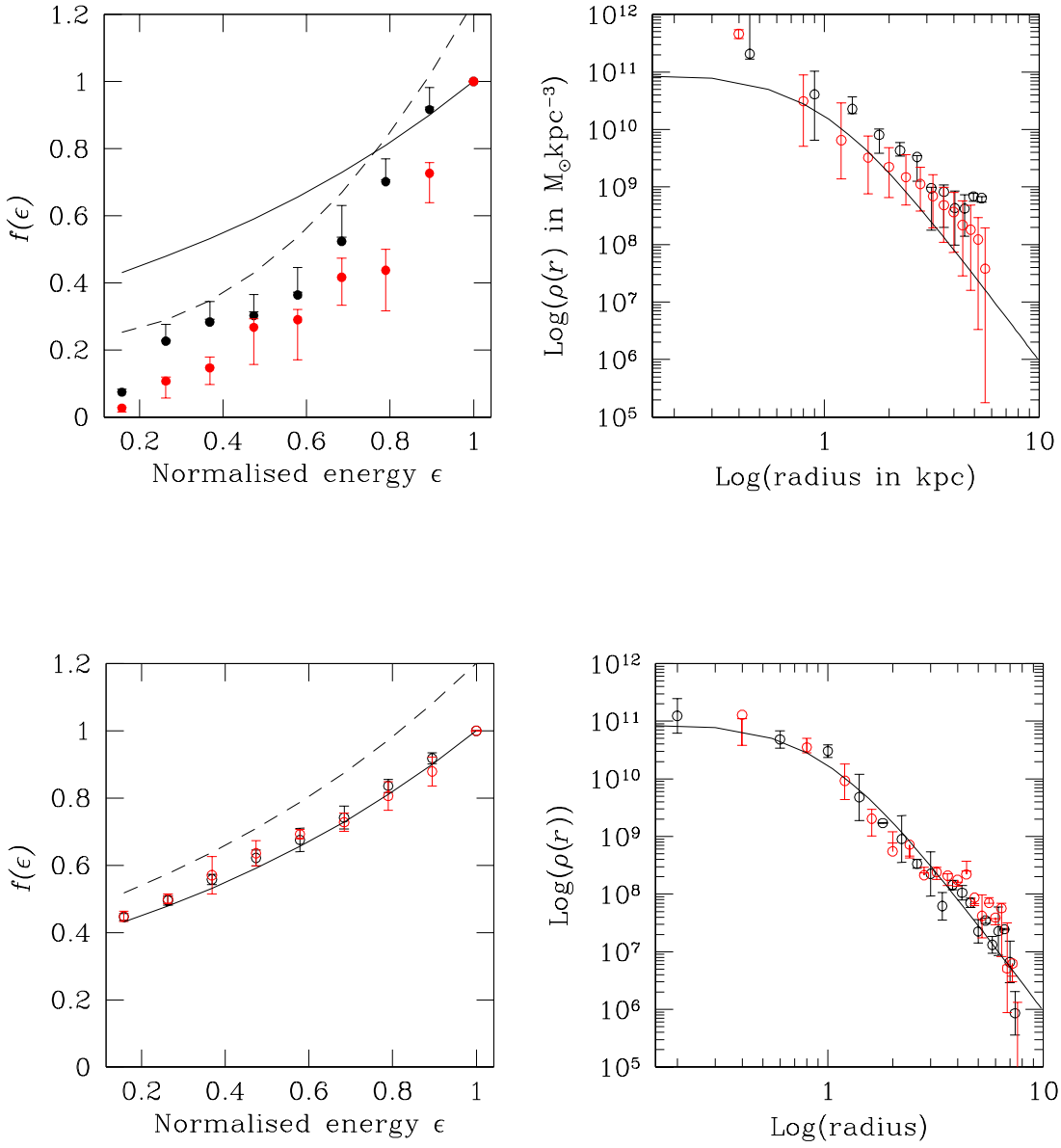

FIG 4. *Figure showing the phase space density vector* $\mathbf{f}^{(i)}$ *(left panels) and gravitational mass density vector* $\boldsymbol{\rho}^{(i)}$ *(right panels) learnt using the synthetic data* $\mathbf{D}_{iso}$*) (lower panels) that is sampled from a known isotropic phase space density and* $\mathbf{D}_{aniso}$*) (upper panel) that is sampled from a chosen anisotropic phase space density. The true gravitational mass density is presented in the solid black lime in the left panels. The true isotropic phase space density that* $\mathbf{D}_{iso}$ *is sampled from, is shown in the lower left panel in the black solid line. The true anisotropic phase space density for the value of the variable* $\mathbf{L} := \mathbf{x} \wedge \mathbf{v} = 0$ *is depicted in black solid line in the upper left panel. The figure also shows in red, the posterior maximising system parameter vectors* $\mathbf{f}^{(i\star)}$ *and* $\boldsymbol{\rho}^{(i\star)}$*, learnt using null-abiding data that are (rejection) sampled from the modal* $\mathbf{f}^{(Mi)}$ *(using* $\boldsymbol{\rho}^{(Mi)}$*) that is itself learnt using synthetic data* $\mathbf{D}_i$*. Here* i *stands for "*iso*" or "*$aniso$*". Modal values of all learnt parameters (components of learnt vectors) are represented by open circles while the 95% HPD credible region is shown by the error bar.*

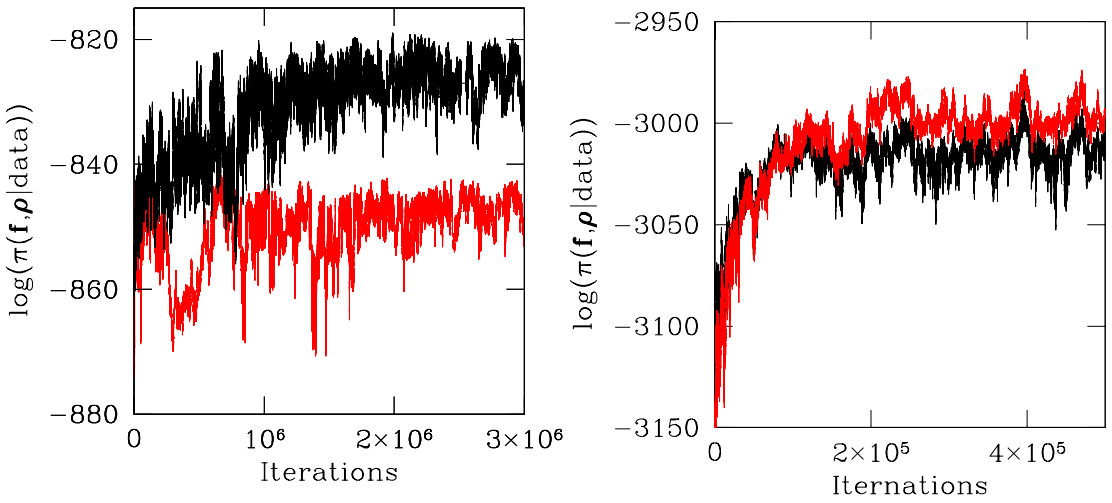

FIG 5. *Figure showing log of the joint posterior probability density* $\pi(\mathbf{f}^{(iso)}, \boldsymbol{\rho}^{(iso)}|\mathbf{D}_{iso})$ *(right) and* $\pi(\mathbf{f}^{(aniso)}, \boldsymbol{\rho}^{(aniso)}|\mathbf{D}_{aniso})$ *(left), in black, for chains that were run for* $5{\times}10^5$ *and* $3{\times}10^6$ *iterations respectively. The log of the joint posterior probability computed using the generated data* $\mathbf{D}_{iso}^{(gen)}$ *and* $\mathbf{D}_{aniso}^{(gen)}$*, are shown in red in the right and left panels respectively. Here, the posterior under the null is maximised during chains run with the null-abiding data* $\mathbf{D}_{iso}^{(gen)}$ *and* $\mathbf{D}_{aniso}^{(gen)}$*, generated by sampling from modal model parameter vectors that are recovered using the data* $\mathbf{D}_{iso}$ *and* $\mathbf{D}_{aniso}$*, respectively. Here* $\mathbf{D}_{iso}$ *has a sample size of 270 and* $\mathbf{D}_{aniso}$ *bears information of 54 particles.*

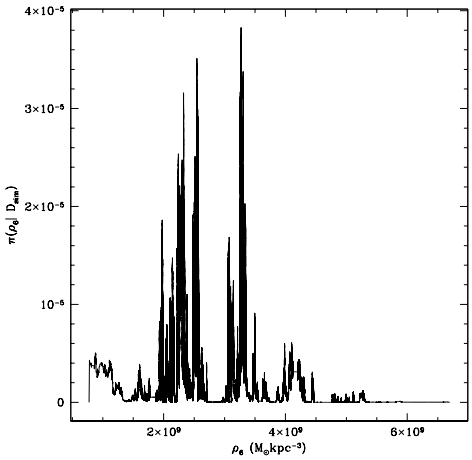

FIG 6. *Marginal posterior probability density of the 6-th component of the* $\boldsymbol{\rho}$ *vector leant using simulated data* $\mathbf{D}_{aniso}$*.*

TABLE 3
Table showing support in data $\mathbf{D}_i$ for null H_i, $i = 1, 2$, computed using different chains.

Chain name	Data set used	$\Pr(H_i|\mathbf{D}_i)$
$PNe - RUN\ I$	$\mathbf{D}_1$	0.61
$PNe - RUN\ II$	$\mathbf{D}_1$	0.58
$PNe - RUN\ III$	$\mathbf{D}_1$	0.62
$GC - RUN\ I$	$\mathbf{D}_2$	0.96
$GC - RUN\ II$	$\mathbf{D}_2$	0.96
$GC - RUN\ III$	$\mathbf{D}_2$	0.93

Thus, $f_1(\boldsymbol{x}, \boldsymbol{v}) \neq f_2(\boldsymbol{x}, \boldsymbol{v}) \Longrightarrow \mathcal{V}_1 \neq \mathcal{V}_2$ where $f_1(\boldsymbol{x}, \boldsymbol{v})$ is phase space pdf defined in volume $\mathcal{V}_1 \subset \mathcal{W}$ and $f_2(\boldsymbol{x}, \boldsymbol{v})$ is phase space pdf defined in volume $\mathcal{V}_2 \subset \mathcal{W}$. In terms of the phase space structure of this real galaxy NGC 3379, we can then conclude that the phase space of the system is marked by at least two distinct volumes, motions in which do not communicate with each other, leading to distinct orbital distributions being set up in these two volumes, which in turn manifests into distinct pdfs for these subspaces of the galactic phase space. The PNe and GC samples are drawn from such distinct pdfs.

Of course, such an interpretation would hold up if we can rule out extraneous reasons that might be invoked to explain the differential support in $\mathbf{D}_1$ and $\mathbf{D}_2$ to the assumption of phase space isotropy. Such extraneous factors are systematically dealt with in Section 8.

It merits mention that our result that $\Pr(H_2|\mathbf{D}_2) \approx 1$ also suggests that the phase space density that the observed GCs in this galaxy live in, is nearly isotropic.

7.1. *The estimate for the whole galaxy.* The motivating idea is that if different data sets are sampled from different volumes of the galactic phase space, where there is no communication amongst these volumes, the gravitational mass density estimates obtained using these data sets will be different. None of these individual estimates will however tell us of the gravitational mass density function of the whole galaxy, in general. The worrying implication of this is that interpreting one of these estimates of $\rho(\mathbf{x})$ as the galactic estimate, can be completely erroneous. The estimate achieved given one data set will then reveal no more about the entire galaxy than the very isolated phase space volume from which these data on particle motions are sampled.

Ideally speaking, the gravitational mass density of the galaxy, at any $|\mathbf{x}|$ can be is approximated as $\max\{\rho_1(x), \rho_2(x), \ldots, \rho_n(x)\}$ where $\rho_i(x)$ is the estimate obtained using the i-th data set, $i = 1, \ldots, n$, where there are n available data sets that have been drawn from n mutually insulated volumes or sub-spaces $\{\mathcal{V}_i\}_{i=1}^n$ within the galactic phase space, namely from $\mathcal{V}_1 \subset \mathcal{W}, \mathcal{V}_2 \subset \mathcal{W}, \ldots, \mathcal{V}_n \subset \mathcal{W}$. The more the number of such data sets available, i.e. bigger is n, the better is the approximation suggested above. However, in this framework we are always running the risk of misidentifying the estimate of a property of a subspace of the galaxy as that of the galactic property.

Here, the i-th data set $\mathbf{D}_i \sim f_i(\mathbf{x}, \mathbf{v})$ where $f_i(\mathbf{x}, \mathbf{v})$ is the pdf that describes sub-space $\mathcal{V}_i \subset \mathcal{W}$. Then the statement that $\mathcal{V}_i$ and $\mathcal{V}_j$ are "mutually insulated" implies that motions in $\mathcal{V}_i$ do not cross over into $\mathcal{V}_j$ and vice versa, resulting in $f_i(\mathbf{x}, \mathbf{v}) \neq f_j(\mathbf{x}, \mathbf{v})$, $j = 1, 2, , \ldots, n$. Thus, $\mathbf{D}_1, \mathbf{D}_2, \ldots, bfD_n$ are not identically distributed. Hence collating all data sets $\{\mathbf{D}_i\}_{i=1}^n$, as a single input into any method that attempts learning the galactic gravitational mass density by invoking phase space densities, will not work, if the method demands the data to be iid. (All methods of learning $\rho(\mathbf{x})$ are underlined by the need to invoke the phase space pdf). Still, if we are in possession of knowledge of the correlation

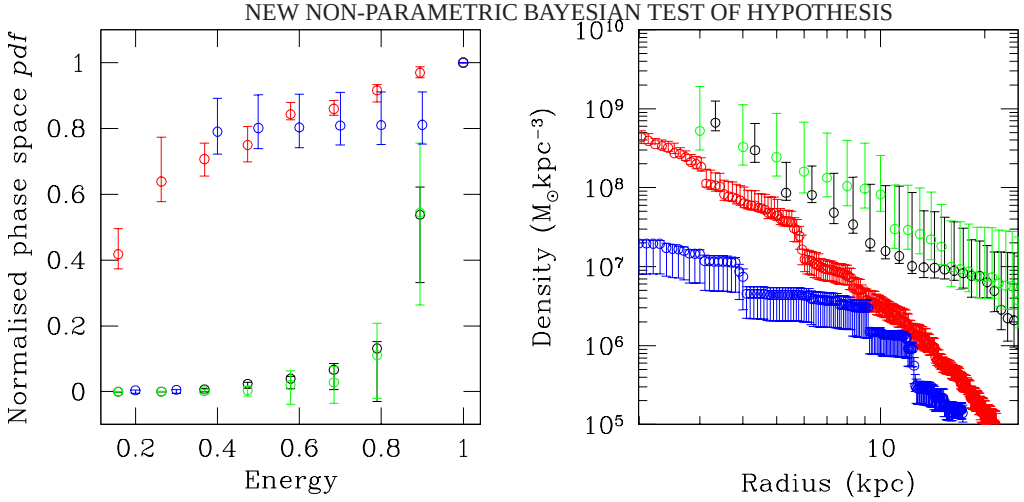

FIG 7. *Right panel: gravitational mass density vector $\boldsymbol{\rho}^{(2)}$ (in black) recovered from chain $GC-RUN\ I$ run using data $\mathbf{D}_2$ and $\boldsymbol{\rho}^{(1)}$ from chain $PNe-RUN\ I$ run using $\mathbf{D}_1$ (in red). These gravitational mass density results were obtained under the assumption of an isotropic phase space, the support for which in the two data sets is indicated in Table 2. Overlaid on these are the identified vectors $\boldsymbol{\rho}^{(1\star)}$ (in blue) and $\boldsymbol{\rho}^{(2\star)}$ (in green), which are respectively, the posterior-maximising gravitational mass density vectors identified in chains run with data $\mathbf{D}_1^{(gen)}$ and $\mathbf{D}_2^{(gen)}$. These data sets are (rejection) sampled from the learnt modal phase space density vectors $\mathbf{f}^{(M1)}$ and $\mathbf{f}^{(M2)}$ respectively. The concurrence of $\boldsymbol{\rho}^{(2)}$ and $\boldsymbol{\rho}^{(2\star)}$ is noted, along with the lack of consistency between $\boldsymbol{\rho}^{(1)}$ and $\boldsymbol{\rho}^{(1\star)}$. In the left panel, the phase space density vectors $\mathbf{f}^{(1)}$ (in red) and $\mathbf{f}^{(2)}$ (in black), learnt from the chains $PNe-RUN\ I$ and $GC-RUN\ I$, are shown, compared respectively to $\mathbf{f}^{(1\star)}$ (in blue) and $\mathbf{f}^{(2\star)}$ (in green). Again, the overlap of $\mathbf{f}^{(2)}$ and $\mathbf{f}^{(2\star)}$ is noted, as is the discord between $\mathbf{f}^{(1)}$ and $\mathbf{f}^{(1\star)}$, especially at higher energies. The $\mathbf{f}$ vectors are normalised to unity at $\epsilon = 1$.*

between the pdfs $f_i(\mathbf{x},\mathbf{v})$ and $f_j(\mathbf{x},\mathbf{v})$, we could use the collated data set in such a method to learn the galactic mass density. However, as there is absolutely no measured, simulated or theoretical information available on the correlation structure between mutually insulated sub-volumes inside galaxies, such a modelling strategy is untenable. Indeed, the development of any such learning strategy, can imply all n available data sets jointly, but will be highly sensitive to the non-linear dynamical modelling of the phase space of the particular galaxy in question. At this stage we recall that for the majority of galaxies, n=1 and n is at most 2 for a few galaxies; examples of these systems include NGC 3379 (Bergond et al., 2006; Douglas et al., 2007), CenA (Woodley & Chakrabarty, in preparation).

Thus, we can put a lower bound on the total gravitational mass content of the galaxy, as we now present for NGC 3379. NGC 3379 is advanced as a dark matter rich galaxy, with the gravitational mass inside a radius of about 20 kpc to be at least as high as about 4 to $10\times10^{11}\mathrm{M}_\odot$, where $\mathrm{M}_\odot$ denotes the mass of the Sun and the astronomical unit of length, kiloparsec, is abbreviated as kpc.

8. Discussions. In the above test, a high support in the GC sample towards an isotropic phase space pdf, along with a moderate support in the sampled PNe for the same assumption, indicate that the two samples are drawn from two distinct phase space density functions.

The expectation that the implementation of the PNe and GC data sets will lead to concurring gravitational mass density estimates is foreshadowed by the assumption that both data sets are sampled from the same - namely, the galactic - phase space density $f(\mathbf{x},\mathbf{v})$. The apparent motivation behind this assumption is that since both samples live in the galactic phase space $\mathcal{W}$, they are expected to be sampled

from the same galactic phase space density, at the galactic gravitational potential. However, such does not necessarily follow if - for example - $f(\mathbf{x},\mathbf{v})$ is a non-analytic function:

$$f(\mathbf{x},\mathbf{v}) = f_p(\mathbf{x},\mathbf{v}), \quad \forall(\mathbf{x}^T,\mathbf{v}^T)^T \in \mathcal{V}_p \subseteq \mathcal{W}, \quad p = 1,\ldots,p_{max} \tag{8.1}$$

Then, if the GC data $\mathbf{D}_2$ are sampled from the density $f_\zeta(\cdot,\cdot)$ and PNe data $\mathbf{D}_1 \sim f_\eta(\cdot,\cdot)$, where $\eta,\zeta \in \{1,\ldots,p_{max}\}$, $\eta \neq \zeta$, then the statement that both the observed samples are drawn from equal phase space densities is erroneous. If this assumption is erroneous, i.e. if observed data $\mathbf{D}_1$ and $\mathbf{D}_2$ are sampled from unequal phase space pdfs,

- firstly it implies that the phase portrait of the galaxy NGC 3379 is split into at least two sub-spaces such that the observed GCs live in a sub-space $\mathcal{V}_\zeta \subseteq \mathcal{W}$ and the observed PNe live in a distinct sub-space $\mathcal{V}_\eta \subseteq \mathcal{W}$, $\eta \neq \zeta$.
- secondly that the phase space densities that describe $\mathcal{V}_\eta$ and $\mathcal{V}_\zeta$ are unequal.

Qualitatively we understand that if the galactic phase space $\mathcal{W}$ is split into isolated volumes, such that the motions in these volumes do not mix and are therefore distinctly distributed in general, the phase space densities of these volumes would be unequal. This is synonymous to saying that $\mathcal{W}$ is marked by at least two distinct basins of attraction and the two observed samples reside in such distinct basins.

Thus, a split $\mathcal{W}$ will readily explain lack of consistency in the estimate of support in the two data sets for the null that the phase space density function that these two data sets are drawn from are isotropic. However, the fundamental question is really about the inverse of this statement. Does differential support in $\mathbf{D}_1$ and $\mathbf{D}_2$ to isotropic pase space pdfs necessarily imply a split $\mathcal{W}$? Indeed it does, since differential support in $\mathbf{D}_1$ and $\mathbf{D}_2$ towards isotropy of the pdf that describes the respective native phase space $\mathcal{V}_1 \subseteq \mathcal{W}$ and $\mathcal{V}_2 \subseteq \mathcal{W}$, implies that the distribution of the phase space vector $(\mathbf{x}^T,\mathbf{v}^T)^T$ in the sub-spaces $\mathcal{V}_1$ and $\mathcal{V}_2$ are distinct. This in turn implies that the phase portrait of this galaxy manifests at least two volumes, the motions in which are isolated from each other. Such separation of the motions is possible if these distinct sub-spaces that the two datasets reside in, are separated by separatrices (Thompson and Stewart, 2001) into isolated volumes, motions across which do not mix.

As for a physical reason for the phase space distribution of the PNe to be less isotropic than the GC population, we can only speculate at the level of this paper. The GC population being an older component of a galaxy (than the PNe population), might have had longer time to equilibrate towards an isotropic distribution. Also, PNe being end states of stars, the PNe population is likely to reside in a flatter component inside the galaxy, than the GC population.

8.1. *What could cause a split galactic phase space.* Galactic phase spaces can be split given that a galaxy is expectedly a complex system, built of multiple components with independent evolutionary histories and distinct dynamical timescales. As an example, at least in the neighbourhood of the Sun, the phase space structure of the Milky Way is highly multi-modal and the ensuing dynamics is highly non-linear, marked by significant chaoticity. The standard causes for the splitting of $\mathcal{W}$ include the development of basins of attraction leading to attractors, generated in a multistable galactic gravitational potential. Basins of attraction could also be triggered around chaotic attractors, which in turn could be due to resonance interaction with external perturbers or due to merging events in the evolutionary history of the galaxy.

8.2. *Assuming spherical geometry.* One worry that astronomers have expressed in the literature about the galaxy NGC 3379 is that the spatial geometry of the system is triaxial and not spherical. For

us, the relevant question to ask is if there is support in the data for the consideration of the gravitational mass density to depend on the spherical radius r alone. In our work, the fact that the methodology assumes the gravitational mass density to be dependent on r^2:=$\mathbf{x}\cdot\mathbf{x}(= x_1^2 + x_2^2 + x_3^2)$ is a manifestation of the broader assumption of phase space isotropy (see Section 3.1). Thus, our test for phase space isotropy includes testing for the assumption that the gravitational mass density of the galaxy NGC 3379 bears a dependence on the spatial coordinates x_j, only via r, $j = 1, 2, 3$. In other words, we have already tested for the assumption that the gravitational mass density depends on the components of the spatial vector, via the spherical radius r.

8.3. *True Mass Distribution.* In contrast to the PNe and GC samples observed in this galaxy, if two observed sets of galactic particles can be inferred to have been drawn from the same phase space density, we will expect consistency in the gravitational matter density that is recovered by using such data sets in a mass determination formalism. At the end of the discussion presented above, we will naturally want to know what the true gravitational mass density function of NGC 3379. However, if distinct density estimates are available at a given radius, from n independent data sets as $\{\rho_1(r), \rho_2(r), \ldots,\ \rho_n(r)\}$, then the lower limit on the galactic gravitational mass density at r is $\sup\{\rho_1(r), \rho_2(r), \ldots,\ \rho_n(r)\}$.

8.4. *Risk of using particle kinematics.* The above results and arguments suggest that it is inherently risky to refer to the gravitational mass density recovered using an observed particle sample - and the gravitational potential computed therefrom - as the gravitational potential of the galaxy. We have demonstrated this with the example of NGC 3379 and shown that inconsistencies in $\rho(r)$ learnt using distinct observed samples, cannot be attributed to any other factor except that these observed samples are drawn from distinct and insular sub-spaces within the galactic phase space, and/or the lack of time-independence in the gravitational mass density or phase space density function.

SUPPLEMENTARY MATERIAL

Supplement A: Motion traces gravitational field - the Earth-Sun system
(). That motion tracks gravitational field is not entirely alien to our experience - after all, the revolution of the Earth around the Sun, at a speed of $V_\oplus$ and radius $R_\oplus$, allows for an estimate of the solar gravitational mass $M_\odot$ via Newton's law, as $V_\oplus^2 = GM_\odot/R_\oplus$, where G is a known constant and the assumption of a circular orbit is made. In contrast to this, when the topology of the orbits in phase space is unknown, a closed-form relation between the orbital speed and gravitational mass enclosed within the orbit cannot be written. Then, in the absence of information about the topology of the orbits of galactic particles, a parametric approximation for the probability density function of $\mathcal{W}$ maybe too naive a model if the system manifests even moderate complexity.

Supplement B: Comparison with the conventional point of view
(). To sum up the results obtained above, we state that in general, kinematic data drawn from distinct phase space densities will yield distinct gravitational mass density functions and this is not solely because of the lack of sufficient information to help constrain the state of anisotropy in the phase space of the system. Here we illustrate the possibility that even when phase space anisotropy parameters - as defined in astronomical literature - are equal given unequal phase space pdfs, the learnt gravitational mass estimates will be in general be unequal. On the very outset this is not surprising at all - after all, equal anisotropy parameters suggest equal ratios of the second order moments of the phase space pdf, which is of course not suggestive of equal phase space pdfs in general. The illustration is performed using the method that is conventionally used to learn gravitational mass enclosed within radius

r, $M(r) = \int_0^r \rho(s) 4\pi s^2 ds$, is the Jeans equation that implements the (typically smoothed) empirical dispersion σ_3 of the values of V_3 of the galactic particles in an observed sample, and the number density function of the sampled particles, in the assumed spherical spatial geometry, i.e. the number of particles $\nu(r)$ that lie in the interval $(r, r+\delta r]$ where δr is the width of the radial bin chosen by the astronomer. Jeans equation gives

$$M(r) = -\frac{r\sigma_3^2}{G}\left[\frac{d\ln\nu(r)}{d\ln r} + \frac{d\ln\sigma_3^2(r)}{d\ln r} + \beta(r)\right] \tag{8.2}$$

where the anisotropy parameter is defined as $\beta(r) = 1 - \sigma_2^2/\sigma^3$. (Since $\beta(\cdot)$ is a function of r, we refer to it as the anisotropy function rather than the anisotropy parameter from now). Then we see that it is possible to obtain distinct $M(r)$ using two data sets that are drawn from two distinct phase space distributions that are equally anisotropic (as parameterised by the anisotropy parameter $\beta(r)$) - then $d\ln\sigma_3(r)/d\ln r$ and $d\ln\nu(r)/d\ln r$ terms would in general be different in the two cases, even if $\beta(r)$ is the same. This follows from the fact that the second moments (leading to the dispersion $\sigma_3(r)$) and the zeroth moments (leading to $\nu(r)$) of unequal phase space density functions are unequal in general, even if the anisotropy functions are equal. Thus, for example, if two observed samples are drawn from phase space densities, each of which is isotropic, but are unequal functions, the anisotropy functions will be identically zero in each case, but the second and zeroth order moments would not necessarily enjoy equal rate of change with $\ln r$, i.e. the $M(r)$ learnt using the two samples will be inconsistent even in this case. Thus, unequal gravitational mass density estimates necessarily imply distinct phase space density functions that the observed data are sampled from, but do not necessarily imply unequal anisotropy functions.

Acknowledgments. I am thankful to Dr. John Aston for his kind comments. I acknowledge the support of a Warwick Centre for Analytical Sciences Fellowship.

References.

AITKIN, M. (1991). Posterior Bayes factors. *Journal of the Royal Statistical Society Series B* **53** 111-142.

BARBIERI, M. and BERGER, J. (2004). Optimal predictive model selection. *The Annals of Statistics* **32** 870-897.

BASU, D. (1975). Statistical Information and Likelihood. *Sankhya A* **37** 1-71.

BELL, E. F. and DE JONG, R. S. (2001). Stellar Mass-to-Light Ratios and the Tully-Fisher Relation. *Astrophysical Journal* **550** 212.

BERGER, J. and PERICCHI, L. (1996a). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association* **57** 109-122.

BERGER, J. and PERICCHI, L. (1996b). The intrinsic Bayes factor for linear models. In *Model Selection, IMS Lecture Series - Monograph Series* (J. BERNARDO, J. BERGER, A. DAWID and A. S. (EDS.), eds.) **5** 25-44. Oxford University Press.

BERGER, J. O. and PERICCHI, L. R. (2001). Objective Bayesian Methods for Model Selection: Introduction and Comparison. In *Model Selection, IMS Lecture Series - Monograph Series* (P. L. (ED.), ed.) **38** 135-207. Beachwood, OH: Institute of Mathematical Statistics.

BERGER, J. and PERICCHI, L. (2004). Training samples in objective Bayesian model selection. *The Annals of Statistics* **32** 841-869.

BERGER, J. and WOLPERT, R. L. (208). The likelihood principle: a review, generalizations, and statistical implications. In *IMS Lecture Notes–Monograph Series* (S. S. G. (ED.), ed.) **6** 25-44. Hayward, CA: Institute of Mathematical Statistics.

BERGOND, G., ZEPF, S. E., ROMANOWSKY, A. J., SHARPLES, R. M. and RHODE, K. L. (2006). Wide-field kinematics of globular clusters in the Leo I group. *Astronomy & Astrophysics* **448** 155-164.

BINNEY, J. and TREMAINE, S. (1987). *Galactic Dynamics*. Princeton University Press, Princeton.

BIRNBAUM, A. (1962). On the foundations of statistical inference. *Jl. of the American Statistical Association* **57** 269-326.

CANO, J. A. and SALMERON, D. (2013). Integral Priors and Constrained Imaginary Training Samples for Nested and Non-nested Bayesian Model Comparison. *Bayesian Analysis* **8, 1** 1–128.

CASELLA, G., GIRN, F. J., MARTNEZ, M. L. and MORENO, E. (2009a). Consistency of Bayesian procedures for variable selection. *Annals of Statistics* **37** 1207-122.

CASELLA, G., GIRN, F. J., MARTNEZ, M. L., , and MORENO, E. (2009b). Consistency of Bayesian procedures for variable selection. *Annals of Statistics* **37, 3** 1207–1228.

CHAKRABARTY, D. (2006). An Inverse Look at the Center of M15. *Astronomical Jl.* **131** 2561-2570.

CHAKRABARTY, D. and PORTEGIES ZWART, S. (2004). An Inverse-Problem Approach to Cluster Dynamics. *Astronomical Jl.* **128** 1046-1057.

CHAKRABARTY, D. and RAYCHAUDHURY, S. (2008). The Distribution of Dark Matter in the Halo of the Early-Type Galaxy NGC 4636. *Astronomical Journal* **135** 2350-2357.

CHIB, S. and JELIAZKOV, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association* **96, 453** 270–281.

CHIPMAN, H., GEORGE, E. and MCCULLOCH, R. E. (2001). The practical implementation of Bayesian model selection (with discussion). In *Model Selection, IMS Lecture Series - Monograph Series* (P. L. (ED.), ed.) **38** 67-134. Beachwood, OH: Institute of Mathematical Statistics.

COCCATO, L., GERHARD, ., ARNABOLDI, M. and ET AL., (2009). Kinematic properties of early-type galaxy haloes using planetary nebulae. *Monthly Notices of the Royal Astronomical Society* **394** 1249.

CÔTÉ, P., MCLAUGHLIN, D. E., HANES, D. A., BRIDGES, T. J., GEISLERAND D. MERRITT, D., HESSER, J. E., HARRIS, G. L. H. and LEE, M. G. (2001). Dynamics of the Globular Cluster System Associated with M87 (NGC 4486). II. Analysis. *Astrophysical Jl.* **559** 828-850.

CÔTÉ, P., MCLAUGHLIN, D. E., COHEN, J. G. and BLAKESLEE, J. P. (2003). Dynamics of the Globular Cluster System Associated with M49 (NGC4472): Cluster Orbital Properties and the Distribution of Dark Matter. *Astrophysical Journal* **591** 850.

CHAKRABARTY, D. (2009). CHASSIS - Inverse Modelling of Relaxed Dynamical Systems. In *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation* (R. D. B. ANDERSSEN R. S. and L. T. H. N. (ED.), eds.) 362-368. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation.

DE B. PEREIRA, C. A. and STERN, J. M. (1999). Evidence and credibility: full Bayesian significance test for precise hypotheses. *Entropy* **1** 69-80.

DE B. PEREIRA, C. A., STERN, J. M. and WECHSLER, S. (2008). Can a Significance Test be Genuinely Bayesian? *Bayesian Analysis* **3** 79-100.

DE BLOK, W. J. G., BOSMA, A. and MCGAUGH, S. (2003). Simulating observations of dark matter dominated galaxies: towards the optimal halo profile. *Monthly Notices of the Royal Astronomical Soc* **340** 657-678.

DEKEL, A., STOEHR, F., MAMON, G. A., COX, T. J., NOVAK, G. S. and PRIMACK, J. R. (2005). Lost and found dark matter in elliptical galaxies. *Nature* **437** 707-710.

DOUGLAS, N. G., NAPOLITANO, N. R., ROMANOWSKY, A. J., COCCATO, L., KUIJKEN, K., MERRIFIELD, M. R., ARNABOLDI, M., GERHARD, O., FREEMAN, K. C., MERRETT, H., NOORDERMEER, E. and CAPACCIOLI, M. (2007). The

PN.S Elliptical Galaxy Survey: Data Reduction, Planetary Nebula Catalog, and Basic Dynamics for NGC 3379. *Astrophysical Jl.* **664** 257-276.

FOUSKAKIS, D., NTZOUFRASY, I. and DRAPER, D. (2012). Power-Expected-Posterior Priors for Variable Selection in Gaussian Linear Models Technical Report, University of California, Santa Cruz.

GALLAZZI, A. and BELL, E. F. (2009). Stellar Mass-to-Light Ratios from Galaxy Spectra: How Accurate Can They Be? *Astrophysical Jl. Supplement* **185** 253-272.

GENZEL, R., SCHÖDEL, R., OTT, T., EISENHAUER, F., HOFMANN, R., LEHNERT, M., ECKART, A., ALEXANDER, T., STERNBERG, A., LENZEN, R., CLÉNET, Y., LACOMBE, F., ROUAN, D., RENZINI, A. and TACCONI-GARMAN, L. E. (2003). The Stellar Cusp around the Supermassive Black Hole in the Galactic Center. *Astrophysical Jl.* **594** 812-832.

GHOSH, J. K. and SAMANTA, T. (2001). Model selection - An overview. *Current Statistics* **80** 1135-1144.

GOODMAN, S. (1999). Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine* **130** 1005-1013.

HAARIO, H., LAINE, M., MIRA, A. and SAKSMAN, E. (2006). DRAM: Efficient adaptive MCMC. *Statistics and Computing* **16** 339.

HAN, C. and CARLIN, B. P. (2000). MCMC Methods for Computing Bayes Factors: A Comparative Review. *Journal of the American Statistical Association* **96** 1122–1132.

HAYASHI, E., NAVARRO, J. F. and SPRINGEL, V. (2007). The shape of the gravitational potential in cold dark matter haloes. *Monthly Notices of the Royal Astronomical Society* **377** 50-62.

KASS, R. E. and RAFTERY, A. E. (1995). Bayes Factors. *JASA* **90** 773-795.

KOOPMANS, L. V. E. (2006). Gravitational Lensing & Stellar Dynamics. *EAS Publications Series* **20** 161.

LEONE, F. C., NOTTINGHAM, R. B. and NELSON, L. S. (1961). The Folded Normal Distribution. *Technometrics* **3** 543.

LINK, W. A. and BARKER, R. J. (2006). Model weights and the foundations of multimodal inference. *Ecology* **87** 2626-2635.

LIU, I.-S. (2002). *Continuum Mechanics*. Springer-Verlag, New York.

ŁOKAS, E. L. and MAMON, G. A. (2003). Dark matter distribution in the Coma cluster from galaxy kinematics: breaking the mass-anisotropy degeneracy. *Monthly Notices of the Royal Astronomical Society* **343** 401.

MADRUGA, M. R., PEREIRA, C. A. B. and STERN, J. M. (2003). Bayesian evidence test for precise hypotheses. *Journal of Statistical Planning and Inference* **117** 185-198.

O'HAGAN, A. (1995). Fractional Bayes Factors for Model Comparisons. *Journal of Royal Statistical Society, Ser. B* **57** 99-138.

PELLEGRINI, S. and CIOTTI, L. (2006). Reconciling optical and X-ray mass estimates: the case of the elliptical galaxy NGC3379. *Monthly Notices of the Royal Astronomical Soc* **370** 1797.

PEREZ, J. and BERGER, J. (2002). Expected-posterior prior distributions for model selection. *Biometrika* **89** 491-511.

PIERCE, M. and ET AL., (2006). Gemini/GMOS spectra of globular clusters in the Leo group elliptical NGC 3379. *Monthly Notices of the Royal Astronomical Society* **366** 1253.

ROBERT, C. (2001). *The Bayesian Choice*. Springer, New York.

ROBERTS, M. S. and WHITEHURST, R. N. (1975). The rotation curve and geometry of M31 at large galactocentric distances. *Astrophysical Jl.* **201** 327-346.

ROMANOWSKY, A. J., DOUGLAS, N. G., ARNABOLDI, M., KUIJKEN, K., MERRIFIELD, M. R., NAPOLITANO, N. R., CAPACCIOLI, M. and FREEMAN, K. C. (2003). A Dearth of Dark Matter in Ordinary Elliptical Galaxies. *Science* **301** 1696-1698.

SALUCCI, P. and BURKERT, A. (2000). Dark Matter Scaling Relations. *Astrophysical Jl. Letters* **537** L9-L12.

SOFUE, Y. and RUBIN, V. (2001). Rotation Curves of Spiral Galaxies. *Annual Review of Astronomy & Astrophysics* **39** 137-

174.

THOMPSON, J. M. T. and STEWART, H. B. (2001). *Non-linear dynamics and chaos*. John Wiley & Sons, Chichester.

TRUESDELL, C., NOLL, W. and ANTMAN, S. S. (2004). *The non-linear field theories of mechanics* Volume 3. Springer-Verlag, New York.

WANG, C. C. (1969). On representations for isotropic functions. *Archive for Rational Mechanics and Analysis* **33** 249-267. 10.1007/BF00281278.

WEIJMANS, A. M. and ET AL., (2009). Stellar velocity profiles and line strengths out to four effective radii in the early-type galaxies NGC 3379 and NGC 821. *Monthly Notices of the Royal Astronomical Society* **398** 561.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL, U.K.
d.chakrabarty@warwick.ac.uk
AND DEPARTMENT OF MATHEMATICS
UNIVERSITY OF LEICESTER
LEICESTER LE1 7RH, U.K.
dc252@le.ac.uk